Chapter 2

Suprasyllabic factors in durational variation

2.1 Introduction

Suprasyllabic timing processes associated with constituents and prominences are described here in terms of the domain and locus framework outlined in Chapter 1. Each process suggested by experimental findings is classified as a domain-edge or a domain-span process; cases where such classification is problematic are discussed. The most appropriate characterisation of the domain of each process is considered, whether this might be in terms of constituents—syntactic or prosodic—or in terms of units defined by lexical stresses or phrasal stresses. An attempt is also made to characterise the locus in each case, and it is considered whether similar processes—particularly domain-initial and domain-final processes—with different domains have similar loci. In some cases the evidence currently available is insufficient to resolve these issues and the experimental work presented in Chapters 3 and 4 is intended to shed further light on them: the particular questions to be addressed experimentally are indicated in such cases.

The influence of syntactic constituents upon speech timing is considered in Section 2.2, and reasons are discussed for the occasional failure of syntactic constituency adequately to predict the occurrence of domain-edge durational effects. A theory of prosodic constituency is presented in Section 2.3, which is motivated in part by the shortcomings of syntactic descriptions of domain-edge effects. The timing effects of prominence are considered under two distinct categories, for reasons outlined in Chapter 1: the effect of the distribution of lexical stress is discussed in Section 2.4, and the effect of pitch accent is discussed in Section 2.5.

In some cases, the designs of previous experiments suggest alternative interpreta-

tions from those advanced by the authors. Where questions of experimental design are directly relevant to the experiments presented in later chapters, previous results are reviewed here but critical analysis of experimental design is presented in the introduction to the relevant new experiment. Thus, studies of polysyllabic shortening are evaluated in Chapter 3 and studies of hierarchical domain-edge effects are evaluated in Chapter 4. Other findings are critically evaluated in the present chapter. In particular, the timing consequences of the distribution of lexical stress—although relevant to the general discussion and to experimental design—are not examined experimentally in this dissertation, and so previous work is analysed in detail here.

As the subject of the dissertation is speech timing in English, experiments relating to other languages are not reported as a matter of course, particularly where results between English and other languages are in agreement. Such experiments are discussed where the results pertain to questions that have not been satisfactorily resolved in English. How far the durational effects of constituents and prominences may be universal is beyond the scope of this dissertation, but where such effects may be seen to be phonologically determined, rather than a by-product of articulation, it is likely that cross-linguistic differences will be observed.

2.2 Syntactic constituents

Durational and intonational variation can influence listeners' judgements about syntactic structure, principally through the delimitation of phrases and clauses (for example: Scott 1982; Beach 1991; Price *et al.* 1991; Marslen-Wilson *et al.* 1992; White 1993). Suprasegmental information may serve to indicate the location of a boundary where syntactic structure is ambiguous. White (1993) plays subjects sentence fragments ending in verbs of ambiguous transitivity, for example:

- *When the tractor pulls ...*

and asks them to indicate which of a pair of sentences represents the most likely complete sentence, for example:

1. *When the tractor pulls, the rope is stretched tight.*

2. *When the tractor pulls the rope, the tree crashes down.*

A number of versions of each sentence fragment are prepared by resynthesising real speech, and the only differences between the versions are the duration of the final verb and the fundamental frequency contour on the voiced section of the verb. Subjects are significantly more likely to select Sentence 1, where the verb is clause-final, when the

duration of the verb is greater and when the fundamental frequency contour is a sharp fall-rise rather than a shallow fall.

These experimental manipulations correspond to suprasegmental variations that have been observed at the end of syntactic phrases and clauses in natural speech by, for example, Cooper & Paccia-Cooper (1980) for duration, and Cooper & Sorenson (1981) for intonation. Speakers are not wholly consistent in the production of such boundary cues, however. As Klatt (1976) observes, boundaries such as that between a noun phrase and a verb phrase may or may not be accompanied by lengthening of the pre-boundary syllable: extra-syntactic factors such as the length of the noun phrase appear to influence the occurrence of suprasegmental boundary cues. Shattuck-Hufnagel & Turk (1996) observe that the same sentence may be realised on different occasions as a single prosodically-delimited phrase or as multiple phrases. Speech rate is a contributory factor: more prosodic phrases are realised at slower rates. In some cases syntactic constituent structure is actually violated: suprasegmental boundary cues may indicate that the main boundary in a NP-VP sentence comes after the verb but before the rest of the verb phrase, apparently influenced by a tendency for prosodic phrases to divide utterances symmetrically. Similarly, function words may be prosodically separated from the content words to which they are most closely bound syntactically. An example of the tendencies to ignore the major syntactic boundary in a sentence and to divide an utterance into approximately equal prosodic phrases is provided by the following television announcement[1] , where "|" indicates a phrase boundary):

Newsnight is over |on BBC2.

The major syntactic phrase boundary, which follows the first word, is not realised prosodically, but the compound prepositional phrase *over on BBC2* is prosodically divided, so as to split the whole utterance into two five-syllable phrases.

Segmental phonological processes such as assimilations or elisions have a similarly indirect relationship to syntax, prompting researchers such as Selkirk (1986) and Nespor & Vogel (1986) to postulate hierarchies of prosodic constituents, influenced by syntax but not isomorphic with syntactic constituents. Such prosodic constituents are hypothesised to be the domains of both segmental and suprasegmental processes.

The relevance of theories of prosodic constituency to speech timing effects such as final lengthening is discussed in Section 2.3. It is not certain, however, that all durational processes have an indirect relationship to syntactic structure. In particular, initial lengthening[2], a domain-edge process, and polysyllabic shortening, a domain-span process, may prove to be closely associated with the lexical word. Domain-span

[1]This example was provided by John Wells on the *phonet* mailing list.

[2]This effect is sometimes called "onset lengthening" but this term conflates the position of the effect, the start of some constituent, and its apparent locus, the onset of the initially-placed syllable.

processes have also been associated with higher-level syntactic constituents such as the sentence. In the remainder of this section, the relationship between syntactic constituents and the domains and loci of durational processes is examined, to assess the effectiveness of syntactically-defined domains in predicting the occurrence of such processes.

2.2.1 Domain-edge effects: final lengthening

Speech segments have been proposed to lengthen before a pause, at the end of different types of syntactic phrases, and at the end of lexical words. The experimental evidence for each effect is considered here in terms of a domain-and-locus description of speech timing. As the boundaries of higher-level and lower-level constituents frequently coincide, the domain responsible for observed lengthening is not always unambiguously identifiable; the locus of the effect in each case may serve to distinguish between domains. The limitations of a characterisation of domains in purely syntactic terms are discussed.

Utterance-final lengthening

Klatt (1976:1211) says that a word spoken at the end of an utterance is about the same duration as it would be when spoken in isolation and "perhaps as much as twice as long as it would be at the beginning of a sentence." Researchers finding evidence for this effect generally measure the duration of segments at the end of single sentences read in isolation (for example: Gaitenby 1965; Oller 1973; Klatt 1975), thus the effect is often called prepausal lengthening or sentence-final lengthening. The domain of the effect under these conditions may be loosely termed the "utterance"—a pause-delimited and relatively self-contained stretch of speech—which may frequently correspond with the similarly-named constituent of certain prosodic hierarchies (see Section 2.3 for discussion). The essence of utterance-final lengthening is that it occurs before a significant break in speech, and that it is a large and reliable effect.

This broad definition of the domain of utterance-final lengthening reflects agreement within the literature about the circumstances in which it is observed, but the locus of the effect is less well established. Klatt (1976:1212) says that lengthening "often extends over several syllables," although he does not cite evidence in support of this claim. Similarly, Cummins (1999:476) states, without citing evidence, that utterance-final lengthening is "characterised by global deceleration and a reduction in articulatory effort over several syllables." Klatt, in common with other researchers, identifies the rhyme of the utterance-final syllable as being the position of the most substantial durational increment. Oller (1973) examines the effect using the reiterant nonsense

words *babab*, *bababab*, *babababab* and *bababababab*, in which the position of primary lexical stress is systematically varied. Both the nucleus and coda of utterance-final syllables show some lengthening compared with syllables earlier in the word: the nucleus has a mean duration of 240 ms compared with 140 ms utterance-medially, and the coda has a mean duration of 80 ms compared with 60 ms utterance-medially. There is only a small lengthening effect on the final syllable onset, and only in the case that the syllable is unstressed. Campbell & Isard (1991) support this finding: syllables lengthened sentence-finally show most of the effect in the syllable rhyme. Campbell & Isard's results also suggest that the composition of the rhyme affects the degree of final lengthening of its constituents: open syllables (those without coda consonants) show more lengthening on the vowel nucleus than do closed syllables.

Examination of the reported duration in Oller's study of vowels and consonants in non-final position within reiterant polysyllables does not provide consistent evidence of the progressive lengthening within the word which would indicate a locus of final lengthening extending to the left of the final syllable; some, but not all, results indicate a slight trend in this direction, and the author makes no comment about such results.

In contrast with many studies of speech timing which use reiterant speech, Oller uses a number of different reiterant templates, such as /baɪb/, /baɪbaɪb/ etc.; /sæs/, /sæsæs/ etc.; /stæst/, /stæstæst/ etc.; /bæ/, /bæbæ/ etc.; /pæp/, /pæpæp/ etc. He finds that the results are largely comparable between these different reiterant materials. Utterance-final fricatives, however, show a particularly large amount of final lengthening, greater that found on the vowel: a mean duration of around 210 ms compared with 70 ms non-utterance-finally. The distribution of lengthening within the locus is often not reported in other studies, with durations simply given for the final syllable or the final syllable rhyme. Cambier-Langeveld's (2000) study of Dutch does report subsyllabic durations, indicating that the distribution of phrase-final and utterance-final lengthening tends to be progressive: that is, the final segment shows the greatest effect—both in absolute terms and in proportion to its non-final duration—and the lengthening of preceding segments diminishes with distance from the boundary. Berkovits (1993b) reports a similar finding for the distribution of utterance-final lengthening in Hebrew. For English, Campbell & Isard (1991) report the distribution of sentence-final lengthening for subsyllabic units in terms of z-scores[3] and find that the coda is lengthened slightly more than the nucleus: the former having a z-score of 1.14 and the latter having a z-score of 1.09. More data are required to determine how far this small difference represents a real trend in English.

Certainly, the discrepancy in Oller's results for the various coda consonants in-

[3] Z-scores here represent the difference between the means for particular sentence-final segments and the overall means for the same segments, in terms of the number of standard deviations of the overall mean for each segment.

dicates that the distribution of lengthening within the locus may vary according to segmental identity, an observation which should be considered when assessing other studies which utilise reiterant speech. As Oller points out, the result may be influenced by the difficulty in comparing certain segment durations in medial and absolute-final position, where they may have "differential mechanisms of termination".

A further point may be made about this study which also applies to other experiments investigating final lengthening. The design of the materials makes it very likely that the target words are pitch accented. The durational effects of pitch accent are discussed in Section 2.5, but it should be noted that the presence of an accent could have significant consequences for the interpretation of such results: firstly, the effect may have greater or lesser magnitude than would otherwise have been observed; secondly, the locus of final lengthening may be different in the presence of pitch accent. In Oller's experiment on utterance-medial, phrase-final words discussed below, it may even be the case—in the absence of independent evidence that a phrase boundary is realised—that the observed durational variation is entirely the effect of pitch accent rather than phrase-finality.

Cambier-Langeveld (2000) provides some data for English and Dutch regarding the interaction between utterance position and accent and the distribution of lengthening beyond the utterance-final syllable. She measures the duration of both syllables of the names *Johnny*, *Michael*, *Macy* and *Joseph* in frame sentences of the type:

I think that Joseph told Johnny about Macy.

where the position of each name and the location of focus—and hence nuclear accent—is systematically varied. Table 2.1 shows mean duration of both syllables in utterance-medial and utterance-final words, when accented and unaccented[4]. The data suggest that there is not a strong interaction between utterance-final lengthening and accentual lengthening, particularly for word-final syllables, which show more final lengthening: there the absolute effect is the same for unaccented and accented words, and the proportional effect is less in the accented case because syllables are longer in accented words.

The locus of utterance-final lengthening appears to extend beyond the final syllable according to the data presented in Table 2.1, although the effect is clearly greater closer to the end of the phrase. It may be noted that the utterance-medial words in this comparison are potentially phrase-final and so the magnitude of utterance-final lengthening may be an underestimate of what it would be relative to a context with no final lengthening influence. Furthermore, if phrase-final lengthening is present on the

[4]Cambier-Langeveld (2000) examines the data in terms of the effect of utterance position on the magnitude of accentual lengthening, but the effect of accent on the magnitude of utterance-final lengthening may be derived from the reported means.

	Stressed syllable, word-initial	Unstressed syllable word-final
Unaccented	19.4 (12%)	122.1 (83%)
Accented	5.4 (3%)	121.5 (68%)

Table 2.1: Data taken from Cambier-Langeveld (2000). Mean utterance-final lengthening (ms) of initial and final syllables in words such as *Joseph* and *Macy*, when accented and unaccented. The proportion of lengthening is shown in parentheses.

utterance-medial words and has a less extensive locus than the utterance-final effect, the relative magnitude of the utterance-final effect on the word-initial and word-final syllables may be distorted[5]. It seems clear, however, that some final lengthening may extend to the left of the final syllable where it is unstressed; furthermore, the locus and magnitude of lengthening are not greatly affected by the presence or absence of pitch accent.

Experiment 2, presented in Chapter 4, examines utterance-final lengthening further, to determine how it is distributed within the locus, and to attempt to corroborate Cambier-Langeveld's (2000) finding that lengthening extends beyond the final syllable where the penultimate syllable carries the primary lexical stress, also testing the assertion that lengthening may extend over several syllables. In addition, Experiment 2 looks again for evidence of an interaction between utterance position and accent.

Phrase-final lengthening

Oller (1973) reports an experiment which he says provides evidence of phrase-final lengthening. Using reiterant nonsense words in the template:

The [bab|babab|bababab|...] is on the table

he finds that the vocalic nucleus is substantially longer in the word-final syllable, which is also the final syllable in the noun phrase. In contrast with the utterance-final measurements however, he fails to find much lengthening of coda consonants. In the light of subsequent research, Oller's original interpretation may be re-evaluated, questioning his identification of phrase-final lengthening in this case.

Klatt (1975) identifies the boundary between a noun phrase and a verb phrase as one of the positions at which lengthening of the preboundary syllable may occur, even in the absence of a pause. He also identifies a number of other syntactically-defined locations at which syllables may be lengthened: before conjunctions, between nouns

[5]Data from Cambier-Langeveld (2000) indicate, however, that for Dutch the locus of final lengthening is not more extensive at utterance edges than at intonational phrase edges.

and prepositional phrases and before embedded clauses. Klatt (1976), however, makes it clear that such locations may not always elicit final lengthening effects:

> Goldhor (1976) found that a noun in subject position in a simple sentence is lengthened more when preceded by an adjective than when the noun appears by itself [. . .] Perhaps in a short phrase consisting of a single noun, the speaker does not bother to lengthen the last syllable of the noun to set off the short noun phrase, but rather groups the subject and verb phrase into a single phonological phrase. The addition of an adjective makes it less likely that he will ignore the phrase boundary.(Klatt 1976:1212)

As noted above, Gee & Grosjean (1983) find evidence supporting this view: the boundaries of certain syntactic phrases are not marked by suprasegmental features, and in some cases, the phrasing suggested by such features may contradict the conventional syntactic interpretation. As well as phrase length, Gee & Grosjean suggest that speech rate and symmetry influence the occurrence of boundaries: fewer phrases are manifest at faster speech rates, and there is a tendency to divide utterances up into phrases of approximately equal size.

The approach taken by Cooper & Paccia-Cooper (1980) to this lack of isomorphism between syntactic phrases and suprasegmental features is to postulate an algorithm which takes performance factors such as speech rate and constituent length into account in mediating between syntax and phonetics. A more thoroughgoing approach is to propose a separate phonological structure—the prosodic hierarchy—which is influenced by syntax but distinct from it, as outlined in Section 2.3.

The question of whether a hierarchy of prosodic constituents accurately predicts the occurrence of final lengthening aside, it is clear that syntactic constituents alone do not. This is one reason for questioning the identification by Oller (1973) of phrase-final lengthening in his experiment, where the syntactic phrase comprises only a determiner and a noun: it is uncertain that such a phrase would be delimited by suprasegmental features in normal speech.

Oller's identification of phrase-final lengthening may also be questioned on the grounds that he describes most of the lengthening as occurring on the syllable nucleus, with little consistent evidence of lengthening on the coda. This result contrasts with other studies of phrase-final lengthening which suggest that, as for utterance-final lengthening, the locus of durational variation is the rhyme of the preboundary syllable. Wightman *et al.* (1992) report that both the vocalic nucleus and coda consonants undergo significant preboundary lengthening, with the strongest utterance-medial preboundary effects being comparable to those found utterance-finally[6]. They further suggest that segments preceding the final nucleus do not show preboundary

[6]Wightman *et al.* (1992) also claim that the amount of utterance-medial final lengthening is correlated with the strength of the prosodic boundary, discussed further in Chapter 4.

lengthening, although their method of analysis may mask some smaller variation: in cases where the final syllable is unstressed, they report the normalised duration of the preceding stressed syllable nucleus and the mean normalised duration of all the material intervening between the stressed syllable nucleus and the final syllable nucleus. They found no evidence of preboundary lengthening in this "nucleus-to-nucleus" stretch, and evidence of a slight tendency towards lengthening in the final stressed syllable nucleus when followed by one or more unstressed preboundary syllables.

Two points may be made about these findings. Firstly, the use of the nucleus-to-nucleus measure could mask some localised lengthening within this stretch, such as on the final syllable onset. Secondly, in common with other studies of final lengthening, the effects reported here may be confounded with the effects of pitch accent. This is particularly true of studies of large corpora of speech where the materials have not been designed to separate the two factors. As nuclear pitch accents tend to occur late in an intonational phrase, there will frequently be cases where the preboundary syllable is subject to both final lengthening and accentual lengthening.

As mentioned above, Cambier-Langeveld (2000) finds that accentual lengthening and utterance-final lengthening combine approximately additively in English; in Dutch, she finds that accentual lengthening is markedly attenuated in utterance-final position. With regard to the interaction between accent and utterance-medial phrase-final lengthening, Price *et al.* (1991), analysing the materials used by Wightman *et al.*, report that the lengthening of the vowel nucleus associated with prominence is less if the prominent syllable is part of the last word in an intonational phrase or utterance (the two types of boundaries are grouped together). This suggests a sub-additive interaction between accentual lengthening and final lengthening, such that the effect of both together is less than the sum of the individual effects, although the graphical data presented by Price *et al.* appear support this conclusion only for syllables that are not in word-final position: word-final syllables appear to show at least as large an effect of accent phrase-finality as when phrase-medially, although Price *et al.* make no mention of this apparent trend.

Nakatani *et al.* (1981) record reiterant versions of two-word phrases within frame sentences of normal English words, for example:

We fished in a mama ma we found in the woods

where the reiterant stretch is spoken with the stress pattern of the phrase *remote stream*. They vary the number of syllables and the stress distribution of the two-word reiterant phrases. There is no control over pitch accent placement and examination of the example materials suggests that lexical stress and pitch accent are often confounded. Among the conditions examined is the effect on reiterant syllable duration of posi-

tion in syntactically-defined phrases. They find comparable final lengthening effects for stressed and unstressed syllables: both are lengthened in absolute phrase-final position, by at least 100ms for three out of four speakers (subsyllabic durations are not reported). They further state: "Compared to the elongation of phrase-final syllables, the effects of other phrase positions on syllable duration were very small indeed" (Nakatani *et al.* 1981:97), but report a small lengthening effect on the phrase-penultimate stressed syllable in the case that the phrase-final syllable is unstressed. None of the observed effects are analysed statistically.

It may be assumed that many of the phrase-penultimate stressed syllables are also pitch accented in Nakatani *et al.*'s study, thus the small observed effect may not generalise to syllables that do not receive pitch accent. Of course, it is very likely in normal speech that phrase-final lexical words will carry pitch accent. In a study using real rather than reiterant speech, Turk (1999) reports lengthening of the rhyme, but not the onset, of the primary lexically stressed syllable in phrase-penultimate position, whether the final syllable carries secondary stress (for example, in *Frankfurt*) or is unstressed (for example, in *Boston*). Turk acknowledges that the words measured are pitch accented in both phrase-final and non-final positions: thus, it seems that the locus of phrase-final lengthening may extend to the left of the final syllable, at least in accented context, although lengthening on the absolute final syllable is greater in all cases. Where the primary lexical stress is in absolute final position, there is no evidence of lengthening on the preceding utterance-penultimate syllable: for example, the first syllable in *Tibet* is not lengthened when the word is phrase-final.

The evidence discussed here suggests the locus of final lengthening extends from the rhyme of the final primary stressed syllable to the phrase boundary. The distribution of the effect may be progressive and is slight in all cases on the onset of the final syllable[7]. It remains to be seen whether this description generalises to non-pitch-accented words and to the case where the final primary stressed syllable in the phrase is followed by more than one syllable before the phrase boundary. Although the experiment presented in Chapter 4 is largely concerned with durational effects at the word-level and at the utterance-level, it does provide some data which relate to lengthening at utterance-internal phrase boundaries.

As discussed above, although final lengthening clearly delimits certain utterance-internal phrases, the phrases themselves are not always syntactically defined. The types of phrases that constitute the domain of final lengthening and their relation to syntactic phrases are discussed further in Section 2.3.

[7] As pointed out by Klatt (1976:1213), the distribution of lengthening within the locus also depends on segmental identity (as observed by Oller regarding utterance-final lengthening): "The lengthening seen at a phrase boundary makes phrase-final fricative and sonorant consonants as much as 40–100ms longer than in non-phrase-final word-final positions. Plosive consonants do not seem to lengthen as much at a phrase boundary."

Word-final lengthening

Word-final lengthening is the extra duration that segments receive because of their proximity to the end of a word in the absence of any larger boundary. Research on word-final lengthening has tended to focus on durational variation in stressed syllables in lexical words. Given the uncertainty discussed earlier about the factors that contribute to prosodic boundaries in speech, separation of word boundaries and phrase boundaries has not proved straightforward for researchers. Klatt (1976:1213) observes: "Early investigators reported large word-final lengthening effects (Barnwell 1971; Lehiste 1972; Klatt 1973), but they did not always control for phrase-final lengthening effects. Word-final lengthening has not always been observed by all investigators (Harris & Umeda 1974) and is probably too small an effect to contribute significantly to the decoding of word boundary locations."

There are other durational influences which affect studies of word-final lengthening. The influence of pitch accent must be considered, as outlined above. Furthermore, the effect of within-word position may be confounded with the effect of word size: as discussed in Section 2.2.3, the hypothesised polysyllabic shortening effect proposes an inverse relationship between word size in syllables and stressed syllable duration.

These multiple influences upon stressed syllable duration within words may be one reason for the continuing uncertainty about the existence of a word-final lengthening effect, with commentators even disagreeing on the interpretation of the same experimental results. Discussing a paper by Beckman & Edwards (1990), Fowler (1990:203) says: "Findings in these experiments and elsewhere (Oller 1973; Klatt 1975) suggest to me that word-final lengthening is fairly ubiquitous," whereas Cutler (1990:208) in the same volume and commentating on the same Beckman & Edwards's paper, says: "In contrast to phrase-final lengthening, word-final lengthening occurs inconsistently. It is more evident at a slow rate of speech; even then, not all speakers show it with all sentences."

The experiments of Beckman & Edwards (1990) which show most evidence of word-final lengthening involve the recording by American English speakers of the phrases *pop opposed* and *poppa posed* in carrier sentences designed to elicit appropriate pitch accent and phrase boundaries at three self-selected speech rates. The vowel durations of the first two syllables in the phrase are reported, the results generally suggesting a word-final lengthening effect even in the absence of a following intonational phrase boundary: the first vowel /a/ tends to be longer in the monosyllabic context and the second vowel /ə/ tends to be longer in the disyllabic context. The data are presented subject-by-subject, which makes comparisons between conditions difficult, but the effect appears to be relatively consistent in both the conditions where the word *pop* was accented and where it was unaccented. Word-final lengthening ap-

pears greater for the full vowel than for the schwa and broadly speaking seems more marked at slow speech rate. It is reported that "[the effect] was not significant for all subjects at all rates, but [...] was consistently in the same direction" (Beckman & Edwards 1990:167). A statistical analysis across subjects is not available, so the significance of the mostly small differences observed is uncertain.

Aside from the difficulties in assessing the data due to the method of presentation of the results, there is a specific problem of reliability, alluded to by Cutler (1990:213): the experiment only uses a single phonetic sequence which is manifestly ambiguous in its lexical structure and so there exists "the possibility that the temporal distinctions [the subjects] produced arose from their desire to emphasise the perceived contrast.[8]" Cutler further observes: "Speakers were required to articulate three /p/ phonemes, each followed by a different vowel, in the crucial region to be measured. It is not particularly far-fetched to suggest that the difficulty of articulating such a sequence with clarity might have been a contributory factor in rendering speakers more prone to produce lengthening effects."

These objections are somewhat assuaged in a further experiment reported by Beckman & Edwards (1990), using the phrases *super station*, *superstition* and *Sioux perspective* and measuring the duration of the first two vowels in various accent conditions. The results of this experiment suggest, however, that much of the variation apparent in the other recordings may be artifactual: for four out of six subjects, the only major effect is the greater duration of /u/ in the monosyllable *Sioux*, in the comparison between pitch-accented words. This could be interpreted as evidence of word-final lengthening; alternatively, it could be seen as polysyllabic shortening amplified by pitch accent. Beckman & Edwards do not report onset or coda durations: data about the distribution of the effect within the syllable might help to distinguish word-edge and word-span interpretations, an approach taken in Experiment 2 in Chapter 4. A study by Silverman & Pierrehumbert (1990) of the alignment of accent peaks also reports word-final lengthening effects: for example, the rhyme of the first syllable in *Mom Le Mann* is shorter than in *Mamalie Le Mann*[9], where the syllable carries a pitch accent in each case. Clearly, the polysyllabic shortening interpretation is also available here, and once again the existence of the effect in the absence of pitch accent is uncertain.

Turk & White (1999) look at the effect of word affiliation on the distribution of lengthening due to accent. An unaccented condition suggests no evidence of word-final lengthening or a balance of word-initial and word-final lengthening: in pairs of

[8] As discussed in Chapter 3, Turk & White (1999) use materials with a similar ambiguity of lexical affiliation, and a control condition without this ambiguity reveals no difference in subjects' production of durational variation. In that experiment, however, a large number of different sentences are used and the recordings are made in blocks designed to deflect attention from possible ambiguity.

[9] This is a study of American English, in which the first syllable is homophonous in these two phrases.

phrases such as *bake enforce* and *bacon force*, there is no significant difference in the duration of the central syllable /ən/ where there is no pitch accent in the phrase. A similar lack of difference is observed for central syllables with secondary lexical stress, such as in the phrase pair *there foreclose* and *therefore close*. If the first word in these phrases carries a pitch accent, then the central syllable is longer when affiliated to that word, but the simplest explanation appears to be the blocking of accentual lengthening at a word boundary.

Turk & Shattuck-Hufnagel (2000) look at a number of hypothesised durational effects related to word structure, and suggest that the most consistent interpretation of their data rules out word-final lengthening, particularly in unaccented words. This research is discussed in detail in Chapter 3 in relation to the results of Experiment 1, which tests the existence of word-level timing processes. Experiment 2, presented in Chapter 4, is designed to distinguish the various interpretations of word-level effects, including word-initial and word-final lengthening, accentual lengthening and polysyllabic shortening.

2.2.2 Domain-edge effects: initial lengthening and shortening

Word-initial lengthening

Oller (1973), as outlined in Section 2.2.1, records reiterant nonsense words in sentences such as *Say a bababab*. In none of the experimental contexts are the initial consonants of the target words phrase-initial or utterance-initial. As stated above, it seems likely that the target words are pitch accented although the placement of accent is not mentioned. In his conclusions, he reports that "the initial consonant effect is even more conclusively demonstrated than most final-syllable effects" (Oller 1973:1244), the effect in question being the lengthening of syllable onset consonants in word-initial syllables when compared with word-medial syllables. He finds the effect for a range of consonants: for /b/, based on recordings of seven subjects; for /s/, /p/ and for both consonants in the cluster /st/, based in single-subject recordings. The fact that both the consonants in the onset cluster show lengthening suggests that the locus of the effect is at least as extensive as the syllable onset, rather than localised on the initial consonant of the syllable.

The magnitude of the effect varies according to the lexical stress of the syllable: for example, in unstressed syllables /b/ is reported to be about 30 ms longer in word-initial position than in word-medial position, where its mean duration is about 40–60 ms; in stressed syllables /b/ is about 20 ms longer in word-initial position than in word-medial position, where its mean duration is about 80–90 ms. Clearly, initial lengthening is proportionally much greater for onset consonants in unstressed syllables, as these are shorter but show more lengthening even in absolute terms. This

could be because consonants in stressed syllables are already quite long and do not have much scope for further expansion, for perceptual or for physiological reasons[10].

There is further effect of stress which is not discussed by Oller, but can be inferred from the graphically-presented data. For the consonant /p/, he reports both the duration of the consonant closure and the duration of the whole closure and aspiration interval. From these results, it can be seen that the lengthening of consonant closure in word-initial position is greater for unstressed syllables than for stressed syllables—as for /b/—and also that aspiration duration is greater word-initially for unstressed syllables but greater word-medially for stressed syllables. Estimated from the graph, with a margin of error of about 3 ms either way, aspiration duration is 25 ms word-initially and 15 ms word-medially for unstressed syllables; for stressed syllables it is 25 ms word-initially and 40 ms word-medially.

Oller does not discuss the possibility of the word-initial lengthening effect extending beyond the syllable onset, but there is evidence from the graphical data of a slight trend towards lengthening of the vowel in word-initial syllables in some experiments. Once again, there appears to be an interaction with lexical stress: vowels in word-initial unstressed syllables appear to be longer in many instances than vowels in word-medial unstressed syllables, particularly by comparison with vowels in the second syllable of words. There is no evidence that the locus of word-initial lengthening extends beyond the syllable onset in stressed syllables.

Further evidence for an interaction between stress and word-initial lengthening may be inferred from the results of a study by Cooper (1991), which is primarily concerned with examining the relationship between the place of articulation of voiceless stops and their duration. Two speakers say the nonsense words *pipip*, *titit* and *kikik* in the carrier sentence *Say . . . again*. The primary stress is either on the first or the second syllable, and it may be surmised that the target word is accompanied by a pitch accent, although this is not stated.

The results for the consonant /p/ are shown in Table 2.2. These exemplify the trend for all three consonants, which is similar to the pattern inferred from Oller's study. Closure duration for all the voiceless stops is longer word-initially than word-medially, for both stressed and unstressed syllables. Once again, aspiration duration is longer word-initially only in unstressed syllables: in stressed syllables, aspiration duration is greater word-medially[11]. Comparison between stressed and unstressed syllable onset durations suggests that the source of this difference may be word-initial shortening of aspiration duration in stressed syllables, rather than a word-medial

[10] Voiced stops can only be sustained for a short time before the build up of supraglottal pressure during closure causes voicing to stop.

[11] As in Oller's study, these trends are not discussed but may be estimated from the graphical data: aspiration duration is greater word-medially in all three consonants for one subject, and in two out of three for the other subject.

lengthening. This conclusion may be drawn because the difference in aspiration duration between stressed and unstressed syllables is comparable word-medially to the stressed vs unstressed differences seen for closure duration; word-initially, however, the difference in aspiration duration between stressed and unstressed context is reduced for subject KM and effectively absent for subject ES.

Syllable context		Subject KM		Subject ES	
		Closure	Aspiration	Closure	Aspiration
Stressed syllable	Word-initial	105	54	98	27
	Word-medial	79	65	82	31
Unstressed syllable	Word-initial	83	44	77	33
	Word-medial	54	28	71	29

Table 2.2: Durations (ms) of closure and aspiration for the onset consonant /p/ in the study by Cooper (1991). The durations shown here have been estimated from graphical data and have a margin of error of 1 ms either way.

Given the likely presence of a pitch accent on the target words in Oller's and Cooper's studies, this interaction between syllable stress and word position for aspiration duration may in fact represent an interaction with pitch accent, because the syllable with primary lexical stress in the accented word will carry the accent, at least for nuclear-accented syllables. A study by Fougeron & Keating (1997) using reiterant speech also finds word-initial lengthening effects on onset consonants, but as they only report the duration of the onset consonant /n/ in unstressed syllable context, their study does not shed further light on the possible interaction between syllable stress and word position, particularly its effect on aspiration duration in voiceless stops. Experiment 2, reported in Chapter 4, examines this interaction further, by looking at the duration of a range of onset consonants in stressed syllables which are initial or medial in accented and unaccented words.

The studies reported here do not discuss the possibility that the locus of word-initial lengthening extends beyond the onset consonant, but Turk & Shattuck-Hufnagel (2000) present evidence—discussed in more detail in Chapter 3—that suggests the effect is largely confined to the onset, and this conclusion is supported by evidence from Fougeron & Keating (1997) and Byrd (2000). The locus of initial lengthening is examined further in Experiment 2.

One further point should be made here: studies of word-initial lengthening tend to use reiterant versions of lexical words. It is not clear whether these results generalise to function words, and how they relate to prosodically-defined words. In the latter regard, if the domain of initial lengthening is the prosodic word rather than the lexical word, lengthening may be seen lexical-word-internally in compounds if they comprise

more than one prosodic word. The question of lexical and prosodic word structure in relation to speech timing is considered in more detail in Section 2.3.

Utterance-initial shortening

Another example of a domain-initial shortening effect has occasionally been observed: the shortening of certain onset consonants utterance-initially; that is, immediately following a break in speech. Fourakis & Monahan (1988) find that the unstressed syllable /mə/ has a mean duration of 75ms in an utterance beginning *Machines will* and a mean duration of 107ms in an utterance beginning *Your machines will*. They consider the possibility that this is due to a measurement error, suggesting that the utterance-initial amplitude at the beginning of /mə/ might have been too weak to appear on the spectrogram, but think this explanation is unlikely because the effect is consistent across speakers and because they "would expect that there should be enough energy in the signal four to six [glottal] pulses into the utterance to register on the spectrograph" (Fourakis & Monahan 1988:293).

Fourakis & Monahan do not report subsyllabic durations, so it is impossible to know what the locus of the effect is, but Fougeron & Keating (1997) report onset consonant duration in a study using reiterant speech. For two of their three speakers, utterance-initial /n/ is shorter than word-initial /n/ in utterance-medial position: for example, for one speaker, /n/ is about 55ms long utterance-initially and about 70ms longer utterance-medially. The magnitude of the effect, where observed, is less than that reported for the whole syllable in Fourakis & Monahan's (1988) study, which might indicate that the locus of utterance-initial shortening encompasses the vocalic nucleus as well, but given the small number of subjects used by Fougeron & Keating, this difference may simply be a matter of between-speaker variation. As with the Fourakis & Monahan study, the syllable in question is not the primary stress: in the Fougeron & Keating study, it is the first syllable in a reiterant version of the number *89*. Furthermore, the syllable onsets in both studies are nasal consonants, so it is uncertain whether the effect applies to all types of onset consonants.

Fougeron and Keating's study is concerned in part with discovering whether the magnitude of initial lengthening is correlated with the strength of prosodic boundaries: that is, whether constituents at higher levels of the prosodic hierarchy show greater initial lengthening effects. The design of this and a similar study by Wightman *et al.* (1992) are analysed in detail in Chapter 4. Experiment 2 presented there includes a comparison of word-initial syllable duration in utterance-initial and utterance-medial position, to determine if the utterance-initial shortening of nasal onset consonants in unstressed syllables is observed in stressed syllables as well, and if so, if the effect also applies to other types of onset consonants. The duration of subsequent vowels is

also reported, to determine if the locus of the effect is the syllable onset or something larger.

2.2.3 Domain-span effects: polysyllabic shortening

The polysyllabic shortening effect may be described thus:

The more syllables that there are in a word, the shorter will be the segments within those syllables.

By this definition, it is strictly a domain-span compression effect, as characterised in Chapter 1, affecting the duration of segments within the locus regardless of position. Most investigators (for example: Lehiste 1972; Port 1981) examine the duration of the primary stressed syllable in relation to word length, as it is the only syllable available for comparison between monosyllabic and disyllabic contexts; such studies might be taken to imply that the locus of polysyllabic shortening is the primary stressed syllable. In contrast, Barnwell (1971:87) concludes from his data that "when the number of syllables for words in a given (constant) structural location is increased [...] both the stressed and unstressed vowels decrease in size [and] the consonants also decrease in size although not as much." This assertion suggests that the locus of the effect is co-extensive with the domain: all parts of the word are subject to durational variation due to word length, with other factors—such as, for example, articulatory constraints and perceptual salience—determining how the timing adjustments are worked out.

The question of the locus of polysyllabic shortening may be left aside for the moment, however, as the results of previous studies do not conclusively demonstrate that the effect is attributable to a domain-span process at all. Indeed, studies often cited as evidence for polysyllabic shortening, such as Barnwell (1971), Lehiste (1972) and Klatt (1973), are elsewhere held to demonstrate another process: Klatt (1976), for example, cites these studies as evidence of word-final lengthening, possibly confounded by phrase-final lengthening. As a typical experimental design compares the duration of the primary stressed syllable in words such as *sleep*, *sleepy* and *sleepiness*, the increase in word length coincides with an increase in the distance of the measured syllable from the end of the word, and possibly from the end of a phrase, depending on the structure of the carrier sentence.

There are other potential confounds in previous studies of polysyllabic shortening, such as rhythmical effects and the influence of pitch accent, discussed below. The interpretation of such studies, in particular Lehiste (1972) and Port (1981), is analysed in Chapter 3 and Experiment 1 is designed to eliminate potential confounds to determine if previous observations may indeed be attributed to a word-level process. The results of Experiment 1 are considered in relation to those of Turk & Shattuck-

Hufnagel (2000); Experiment 2, in Chapter 4, attempts to determine the *type* of process occurring at the word-level.

Two other issues should be mentioned at this point. Firstly, as with word-initial lengthening, it is uncertain whether polysyllabic shortening, if it can be demonstrated unambiguously, operates over a syntactic domain—the lexical word—or a phonological domain—the prosodic word or a similar near-word-sized unit. The nature of the prosodic word is discussed further in Section 2.3.2.

Secondly, the study by Huggins (1975) is sometimes cited as evidence for polysyllabic shortening, but this is not his conclusion. He reports the durations of the syllables /tʃiz/ and /baʊnd/ in the context of sentences formed from the template

Cheese(s) (a)bound(ed) (ab)out.

where the presence of the unstressed syllables in brackets is systematically varied. He finds that the duration of /tʃiz/ is reduced by an unstressed syllable within the same word, but not by an unstressed syllable following the word boundary. From this, the existence of a word-level effect might be inferred. He also finds, however, that /baʊnd/ is shortened by an unstressed syllable both within the word and across a word boundary. His conclusion is that the shortening is a rhythmical effect—as discussed in Section 2.5—which is blocked at the major syntactic boundary between *cheese* and *bound*. Furthermore, Huggins (1975:462) states that this conclusion is "speculative, being based on very little data," and Rakerd *et al.* (1987:149) report that Huggins failed to replicate the shortening effect with added sentences and subjects.

2.2.4 Domain-span effects: sentence-level shortening

Lehiste (1974), cited in Klatt (1976:1212), finds "a general tendency for readers to talk faster when phrases or sentences are longer." This effect, also alluded to by Jones (1942–43), suggests a domain-span compression effect, analogous to polysyllabic shortening, but over a larger domain such as the sentence. The domain may also be the phonological utterance, which, as discussed in Section 2.2.1, may generally be equated with the sentence in studies requiring subjects to read isolated sentences. As rate of speech is held to increase over the whole sentence, the locus of such an effect would be co-extensive with the domain. As with polysyllabic shortening, however, it remains uncertain whether this interpretation represents the most accurate account.

In a single subject study, Gaitenby (1965) measures the duration of phrases in sentences of varying lengths.

1. *Why don't you?*

2. *Why don't you get tickets?*

3. *Why don't you get tickets for tomorrow?*

4. *Why don't you get tickets for tomorrow night?*

5. *Why don't you get tickets for tomorrow night's programme?*

She finds that each added phrase is longer in sentence-final position than otherwise, but that, other than this effect, there is no clear pattern of variation due to sentence length. Thus *get tickets* is 740 ms in Sentence 2 and 590 ms in Sentence 3, but Sentence 4 and Sentence 5 cause little further variation in the duration of *get tickets*, which is 590 ms in the former and 580 ms in the latter. Similar results are reported for the other phrases in the sentence.

In contrast, Rakerd *et al.* (1987) find that the duration of a sentence-medial syllable is greater when the sentence in which it is uttered is shorter. They measure the duration of monosyllabic lexical words in a pair of sentences of different lengths, for example, *date* in the context of:

1. *His first date aroused some anxiety.*

2. *His first date aroused some anxiety for obvious reasons.*

The mean duration of the monosyllabic targets is 11 ms greater in shorter sentences such as Sentence 1. This may be seen as evidence of a sentence-span effect, but other interpretations remain available. Firstly, if a domain-span effect is responsible for the observation, the domain may be better characterised as a prosodic entity such as the utterance rather than a syntactic entity such as the sentence. Furthermore, the locus of the effect remains to be established: is it just the duration of lexically-stressed or pitch-accented syllables that is affected by domain length, or does the adjustment take place throughout the domain? If the former, is this because such syllables are structurally significant or simply because they are longer and so have more scope for durational variation?

Secondly, there may be an alternative, indirect explanation of apparent sentence-span shortening: the length of the sentence may influence its phrasing, such that the measured word or syllable stands in a different position with respect to boundaries in the two versions of the experimental sentence. For example, Sentence 1 might be realised as two phrases with a boundary after the subject noun phrase: thus *date* would be subject to phrase-final lengthening. If Sentence 2 elicited a different phrasing—perhaps two phrases with a boundary after the object noun phrase—*date* would be shorter because of its phrase-internal position. Thus an apparent domain-span effect would have a domain-edge explanation.

The experiment described in Chapter 4 is designed to test the existence of domain-span processes in the word and the sentence/utterance.

2.3 Prosodic constituents

As the evidence discussed above demonstrates, syntax does not reliably predict the occurrence of all durational processes associated with constituents of spoken English: phrase-final lengthening, in particular, appears to take place in domains that are only indirectly related to syntactic structure. Prosodic hierarchies have been proposed by researchers such as Nespor & Vogel (1986), Hayes (1989) and Selkirk (1978; 1986; 1996) to account for the domains of occurrence of segmental phonological processes, and in certain cases, intonational phenomena (see Shattuck-Hufnagel & Turk 1996 for a review). It has been proposed by researchers such as Wightman *et al.* (1992) that it is these prosodic constituents which constitute the domains of final lengthening effects. Other researchers, such as Fougeron & Keating (1997), have linked the strength of prosodic constituent boundaries to the magnitude of initial lengthening effects.

Specific details of prosodic hierarchies vary between theories: Nespor & Vogel's (1986) account is presented in Section 2.3.1 as an illustration of how prosodic constituents may be derived. Constituents which may influence durational processes, and which may have different names in different theories, are discussed in Section 2.3.2.

2.3.1 A prosodic hierarchy

Nespor & Vogel's framework is proposed as one of a number of interacting phonological subsystems, "each governed by its own principles, such as the theories of the metrical grid, lexical phonology, autosegmental phonology and prosodic phonology" (Nespor & Vogel 1986:1). This separation reflects one possible standpoint on the disparate influences on speech timing, distinguishing those aspects of durational variation that are directly predictable from syntax; those that can be accounted for within the domains provided by prosodic phonology; those that are better described within a different phonological subsystem, such as metrical phonology; and those that are "better relegated to phonetic interpretation along with other details of phonetic timing such as the different durations of vowels before voiced and unvoiced obstruents" (Beckman & Edwards 1990:158)[12]. Nespor & Vogel's distinction is not always clear-cut, however: they include as evidence for their prosodic constituents metrical effects such as stress shift, and intonational phenomena that might otherwise be regarded as belonging to autosegmental phonology. Some researchers go further and explicitly attempt to link apparently distinct subsystems: for example, Beckman & Edwards (1990,

[12]The influence of syntax on speech timing, originally considered to be direct, is often now regarded as necessarily mediated by prosodic structure. As this discussion in this chapter illustrates, the evidence for this view is stronger for some timing processes, such as phrase-final lengthening, than it is for others, such as the various hypothesised word-level effects.

1994) attempt to identify each level of prominence as the head of a particular prosodic constituent. Whether different influences should be separated or whether they reflect an undiscovered unified representation of speech timing remains to be seen, but some separation appears at present to be necessary: as discussed in Section 2.4, some rhythmical aspects of speech timing do not appear to respect the boundaries of syntactic or prosodic constituents.

Nespor & Vogel present a hierarchy of prosodic constituents—syllable, foot, phonological word, clitic group, phonological phrase, intonational phrase, phonological utterance—which is observable because "each prosodic constituent serves as the domain of application of specific phonological rules and phonetic processes" (Nespor & Vogel 1986:1). The prosodic constituents above the level of the word are constructed according to rules which refer to the output of the morphosyntactic component of the grammar, but also take into account factors such constituent length, branching, and prominence relations.

The phenomena for which prosodic constituents serve as domains include effects such as liaison in French, *raddoppiamento sintattico* in dialects of Italian, nasal assimilation in Spanish, flapping in American English, and linking-r and intrusive-r in dialects of British English. Such processes are often referred to as external sandhi rules, meaning word-external: because the word does not have a privileged position within their framework, Nespor & Vogel do not use the qualifier "external". As mentioned above, they also invoke certain non-segmental phenomena such as iambic reversal (also known as stress shift) in American English as evidence for certain constituents. The application of these processes within the domains delimited by the constituents of the prosodic hierarchy is governed by three types of rules: domain-span rules, domain-limit rules and domain-juncture rules.

Domain-span rules are those that apply within a domain but not at its edge[13]. For instance, the phenomena of linking-r and intrusive-r in certain non-rhotic dialects of British English are accounted for by a phonological utterance-span rule of r-insertion. Thus if any of the vowels /ɔ/, /ɑ/ or /ə/ is followed immediately by another vowel within a phonological utterance, an /ɹ/ may be articulated between them. At utterance boundaries, however, this process is blocked. Iambic reversal is also cast as a domain-span rule of the phonological phrase: an iambic word may be realised as a trochee when followed by a stressed syllable within the same phonological phrase. Thus, although the word *thirteen* is iambic in citation form, in the phrase *thirteen men* it may have its main stress realised word-initially. As a domain-span rule, iambic reversal is blocked by the intervention of a phonological phrase boundary.

[13]The usage of the term "domain-span" is rather different in Nespor & Vogel's account of segmental phonology from its usage in this dissertation, where it refers to durational variation related to the size of a particular constituent and affecting subconstituents regardless of position.

Domain-limit rules apply at the edges of domains but not within domains. An example is the glottalisation of /t/ in English, which may occur when /t/ is syllable-final. The aspiration of /t/ is also described as domain-limit rule, in this case at the level of the foot: /t/ is aspirated in certain dialects of American English when it is foot-initial but not elsewhere.

Domain-juncture rules, like domain-limit rules, occur at the edges of domains but not within them, further requiring that the appropriate conditions are met on the other side of the domain-boundary. The trilling of orthographic *r* in Spanish is described as a domain-juncture rule at the level of the syllable: a syllable-initial *r* may be trilled rather than tapped if the previous syllable ends with a sonorant.

The principles which govern the hierarchical construction of prosodic constituents into trees are in important respects different from those which apply in syntax. In particular, prosodic trees are n-ary branching rather than binary branching, forming an essentially flat structure, rather than binary trees which have in principle unlimited depth. There may be any number of sister nodes at a given level of the prosodic tree: one of them is designated the strong node according to specific principles applying to each level, the others are all designated as weak. The other important principle of construction in Nespor & Vogel's theory of the prosodic hierarchy is the "strict layer rule". This means that any constituent of a given level is exhaustively contained within the superordinate constituent of which it is a part, and conversely, any non-terminal constituent is composed exclusively of one or more immediately lower units. Not all theories of the prosodic hierarchy require that strict layering holds in all circumstances, however: following an optimality theory approach taken by Ito & Mester (1992), Selkirk (1996) analyses the strict layering rule into four constraints called "headedness", "layeredness", "exhaustivity" and "nonrecursivity". In Selkirk's prosodic hierarchy, the latter two constraints are violated in some circumstances: where exhaustivity is violated, a constituent may dominate another constituent of more than one level down; where nonrecursivity is violated, a constituent may dominate another constituent of the same level.

An important characteristic of Nespor & Vogel's approach, particularly in relation to speech timing, is the restructuring of constituents allowed at higher levels of the hierarchy, from the phonological phrase upwards. The length of constituents is important in determining whether larger constituents are formed from the merger of two or more constituents at the same level. In reference to intonational phrase (IP) restructuring, they say that "the faster the rate of speech, the longer the [IPs] of a given utterance tend to be; conversely, the slower the rate of speech, the shorter the [IPs] tend to be. In this way, some more abstract notion of length in terms of timing or

rhythm may be respected"[14] (Nespor & Vogel 1986:195). They further state that "for both [IP] and [phonological utterance] restructuring, what seems to be crucial is some average length in terms of timing" (Nespor & Vogel 1986:240).

Nespor & Vogel's prosodic hierarchy is intended to be universal, with the same set of constituents combining in similar fashion in all languages, although certain parameters—such as the direction of affiliation of lower-level constituents into higher-level ones—may vary between languages. They claim that although a particular language may lack segmental domain rules for some phonological constituents, these constituents may still be necessary in order to define relative prominence relations, or to account for other phenomena such as constituent-final lengthening. This suggestion may be intended to imply that all constituents within the hierarchy, with the likely exception of subword units, manifest final lengthening to some extent. Specific experimental evidence for a hierarchy of constituents delimited by durational effects of increasing magnitude at higher levels—what may be termed "hierarchical lengthening"—is discussed in Chapter 4, where Experiment 2 explores two potential levels of hierarchical variation: the word and the utterance.

2.3.2 Prosodic constituents and speech timing

The general principles of the prosodic hierarchy having been outlined, brief descriptions of particular prosodic constituents are presented in this section. Particular emphasis is given to those constituents that may serve as domains of suprasyllabic durational processes. Points of major divergence between the hierarchy of Nespor & Vogel and other theories, such as that of Selkirk (1996), are also indicated. In accordance with common practice, what Nespor & Vogel call the "phonological word" is referred to here as the "prosodic word". Also, Nespor & Vogel refer to a sub-word constituent which they call the "foot"; as discussed in Section 2.4, there are a number of possible types of feet, and so the term "within-word foot" is used here to avoid ambiguity.

The Syllable

Subsyllabic units such as the onset, rhyme and coda are excluded from Nespor & Vogel's prosodic hierarchy. Firstly, they do not obey the principle of construction which requires each level to be exclusively composed of one type of constituent. Secondly, a number of segmental phonological rules which can be expressed in terms of subsyllabic units can also be expressed in terms of a domain defined by the syllable itself. Syllables are thus the lowest unit in the prosodic hierarchy. Syllabification is according

[14] It may be noted that domain-span compression effects, of the type discussed in Section 2.2.4, make a similar link between constituent size and *local* speech rate, although there the direction of causality is reversed: the rate of speech is faster because the constituent is longer.

the maximal onset principle, over the domain of the phonological word. Resyllabification is possible over larger domains in some languages, such as many European Romance languages, but not in English.

As mentioned in Chapter 1, there are a number of durational processes which may be related to the domain of the syllable. Firstly, the distribution of lexical stress is generally described in terms of syllables: the primary acoustic indicator of lexical stress, together with vowel quality, is lengthening of the stressed vowel and, to a lesser extent, of the onset and coda consonants. Other durational processes relating to the distribution of stressed and unstressed syllables are described in Section 2.4. There are also a number of processes which may be related to the composition of syllables or subsyllabic units, such as the shortening of consonants in clusters and the lengthening of a vowel before a voiced coda.

The Within-Word Foot

Each within-word foot is composed of one relatively strong syllable and a number of relatively weak syllables, up to a prosodic word boundary. In Nespor & Vogel's account, some languages allow only binary feet, but in English any number of weak syllables may be included. Other parameters which vary between languages are the position of the strong syllable—it is the leftmost syllable in English—and the possibility of refooting across words. Nespor and Vogel say that this is not allowed in English: the domain over which feet must be well-formed is the prosodic word.

Studies discussed in Section 2.2.1 have suggested that the within-word foot may be the locus of phrase-final lengthening (for example: Nakatani *et al.* 1981), although such lengthening seems to exclude the onset consonant of the first syllable of the foot, and possibly the other onset consonants. Turk (1999) finds, however, that words such as *Woodstock*, containing two within-word feet, showed phrase-final lengthening on the word-initial syllable as well as on the word-final syllable. As described in Section 2.5, the results of Turk & Sawusch (1997) suggest the possibility that the within-word foot may be the locus of accentual lengthening, although subsequent research such as Turk & White (1999) shows that accentual lengthening can extend beyond the within-word foot. Thus, it now appears that the within-word foot is not the locus of final lengthening or accentual lengthening.

The Prosodic Word

Nespor & Vogel's (1986) rules for the construction of the prosodic word differ between languages—indeed, they say that it is the constituent which varies most between languages—but there are no instances in which a prosodic word may be larger than a terminal node of the syntactic tree. In some languages, such as Greek and Latin,

there is exact identity between these terminal nodes and prosodic words, but for most languages each morphological stem must belong to a separate word, with affixes generally attaching by adjacency to the stem within the same terminal syntactic node. In languages such as Hungarian and Italian, morphological or phonological criteria identify certain affixes which must comprise separate prosodic words. The status of English is not explicitly discussed in Nespor & Vogel's account, although it may be surmised that it belongs to the set of languages identifying each morphological stem with a separate prosodic word.

The prosodic word is constructed from all the feet within its domain. The relative prominence rules require that, within a given language, in the unmarked case the strong foot will consistently be either the leftmost one or the rightmost one in the prosodic word of a given language. In English, it is the rightmost foot which is strong in the unmarked case.

There are two main points of contention regarding the mapping between syntactic and prosodic words: one is the prosodic status of function words, the other is the prosodic nature of compounds.

In Nespor & Vogel's formulation, every syntactic word—lexical and functional—forms a single prosodic word (except where the syntactic word contains more than one morphological stem). This significantly differs from other theories, notably that of Selkirk (1996), which accord prosodic word status to lexical words as a matter of course, but to function words only in certain cases. According to Selkirk, function words form prosodic words in citation form and when in focus (that is, when pitch accented); most also form prosodic words when phonological-phrase-final. In other circumstances, monosyllabic function words are regarded as clitics—unstressed syllables which are not dominated by either a within-word foot or a prosodic word of their own. Disyllabic function words, and a few monosyllabic function words such as *up*, *too* and *off* are regarded as footed by virtue of containing a stressed syllable, but do not form prosodic words of their own. The structures within which function words coordinate with lexical words depend whether they precede a lexical word (proclitics) or follow it (enclitics).

Selkirk (1996) argues that an enclitic function word—for example, the pronoun *him* in *need him*—may join prosodic structure within a nested prosodic word (PW): *((need)*$_{PW}$*him)*$_{PW}$. The verb in *need him* is directly dominated by a prosodic word, which is itself dominated by another prosodic word which also dominates the pronoun; thus, the strict layering constraint of nonrecursivity—described in Section 2.3.1—is violated. A proclitic function word may join prosodic structure at a higher level, such as the phonological phrase (PP); for example, the auxiliary *can* in *can paint*: *(can (paint)*$_{PW}$*)*$_{PP}$; here the strict layering constraint of exhaustivity is violated. In contrast,

Nespor & Vogel (1986) claim both lexical words and function words form prosodic words in their own right; "functional" prosodic words may then combine with "lexical" prosodic words within clitic groups.

Clearly, Nespor & Vogel's theory would predict that prosodic-word-level timing processes should apply to both content words and function words, but Selkirk's account would not, at least for proclitics. The resolution of this issue is not straightforward: given the closed-class nature of function words, and the fact that most are monosyllabic, the experimental examination of processes such as initial lengthening and polysyllabic shortening is rendered highly problematic and to some extent irrelevant. Furthermore, in practical terms, function words tend to be much shorter than lexical words, and thus any systematic durational variation would be slight in, for example, disyllabic function words compared with monosyllabic function words[15]. The issue of the nesting of prosodic words may be more tractable, however: Selkirk says that "words with 'level 2' suffixes [...] in English have exactly the same prosodic structure as verbs with object clitics." Thus, if durational processes apply across a prosodic word domain, they should apply equally to *saw it* and to *sawing*.

The theoretical standpoints regarding compounds are harder to determine, but the main point of contention appears to be whether compounds—words containing two morphological stems, such as *kneecap* and *shakedown*—form single prosodic words, pairs of prosodic words or nested prosodic words. For other languages, apparently including English, Nespor & Vogel suggest that each morphological stem forms its own prosodic word. Selkirk (1984) claims that, syntactically, compounds have a nested structure, for example: $((\text{knee})_W(\text{cap})_W)_W$, and thus both parts of compound are joined within a single word. It is unclear how such a structure is to be interpreted prosodically: as noted above with regard to function words, Selkirk proposes that prosodic words may be nested, so the two prosodic words in a compound may be dominated by a higher prosodic word, but does not explicitly discuss the case of syntactic compounds. She does say, however, that "a sequence of lexical words [...] in morphosyntactic representation [...] is characteristically prosodised as a sequence of prosodic words [...] in phonological representation" (Selkirk 1996:188), which might be taken to imply a one-to-one mapping. As syntactic compound words are nested, they would also be nested prosodically.

If the prosodic word, rather than the lexical word, is the domain word-level processes, then, if the above interpretation of Selkirk's theory is correct, such processes would operate recursively. For example, each prosodic word individually and the

[15] In certain circumstances, such as phrase-finally and when in focus, function words are not reduced: the most obvious timing consequence of this is lengthening. It may be expected that the distribution of such lengthening would reflect that seen in lexical words when phrase-final or accented, but this issue is beyond the scope of this dissertation.

higher prosodic word would exhibit word-initial lengthening and polysyllabic shortening. As mentioned above in relation to function words, the fact that many compounds are formed of two monosyllabic stems makes testing the predictions of the nested and non-nested hypotheses difficult for polysyllabic shortening; for word-initial lengthening, one prediction would be that would be greater in compounds, as a result of the operation of two recursive word-initial processes, than in monomorphemic words: thus, /k/ would be longer on *crowbar* than in *crowing*.

What durational evidence there is on this issue suggests that compounds do exhibit similar durational properties to noncompounds. Barnwell (1971) observes that compounds comprising two monosyllabic words behave for timing purposes as a single disyllabic word. Klatt (1975), based upon a small amount of data, finds that vowel duration is better predicted if compounds are treated as "single phonological words" over which a word-span compression process operates. Turk & White (1999) find similar patterns of accentual lengthening across a range of disyllabic words containing two stressed syllables, including some noncompounds—such as *skimpy* and *capsize*—and compounds—such as *kneecap* and *downstairs*.

The issue of the nesting of prosodic words is not tested directly in Experiment 1 or Experiment 2, and it may be assumed that the term "word" is used to mean a lexical word, whether noncompound or compound.

The Clitic Group

As observed above, clitics are monosyllabic function words that have some phonological dependence on non-clitic constituents. Clitics may not, for example, receive contrastive stress or stand alone as utterances. In Nespor & Vogel's theory, the clitic group dominates single prosodic words and adjacent clitics: the construction of the clitic group makes reference to the syntactic structure of a language, so that in cases where clitics may either encliticise or procliticise, the direction of attachment is often determined by syntactic structure. In most languages, including English, the strong element within the clitic group is the prosodic word that contains the non-clitic element.

The behaviour of clitics in Romance languages such as Italian strongly motivates the hypothesised clitic group, as illustrated by Nespor & Vogel. It is less clear whether languages such as English have phonological processes which support a single level at which all clitics are associated with adjacent lexical words. Selkirk's (1996) treatment of clitics is different: as discussed above, enclitics attach to a prosodic word recursively dominating the preceding prosodic word. Proclitics are not dominated by a prosodic word, but attach to the phonological phrase which dominates the following prosodic word. Thus, in the phrases *a conversion*, *for a massage*, *that you could ask*, all the function

words are unfooted syllables attached to the phonological phrase which dominates the phrase-final prosodic word. Selkirk's phonological phrase may thus be said to share some characteristics of Nespor & Vogel's clitic group[16].

Due, perhaps, to the uncertain status of clitics in English, the timing consequences of this constituent have not been explored, although the formulations of Nespor & Vogel and Selkirk differ in their predictions: if there is a domain-span timing process which applies to the clitic group, proclitics and enclitics should be equivalent in their effect on the size of the constituent according to Nespor & Vogel. In Selkirk's theory however, enclitics should influence prosodic word timing processes, whereas proclitics only affect lexical words though processes which may hold at the phonological phrase level.

The Phonological Phrase

Nespor & Vogel (1986:168) construct the phonological phrase from: "a clitic group which contains a lexical head (X) and all clitic groups on its non-recursive side up to the maximal projection of X." (In languages, such as English, whose syntactic trees are right-branching, the rightmost node of the phonological phrase is the strong element; in languages whose syntactic trees are left-branching, it is the leftmost node which is strong.) For example, the utterance:

The indolent donkey| masticated| the tasty thistle| languorously.

divides as indicated into four phonological phrases

Although the phonological phrase (Φ) has a precise syntactic definition, it is the lowest constituent in the prosodic hierarchy for which optional restructuring is possible, and this operates according to non-syntactic criteria: "A nonbranching Φ which is the first complement of X on its recursive side is joined into the Φ that contains X" (Nespor & Vogel 1986:168). The requirement that the adjacent phonological phrase be nonbranching reflects the intuition that the length of prosodically-defined phrases tends to cluster around some mean, with very short or very longer phrases being less likely. For example, the *indolent donkey* sentence above could be restructured into two phonological phrases by combining the first two phrases and the last two phrases, but this is unlikely because the phrases are already relatively long. A similarly-structured utterance with shorter phrases, such as:

The mouse| chewed| the cheese| fast.

is more likely to undergo restructuring to form two phonological phrases with a boundary following *chewed*.

[16]Selkirk (1996) does not make the distinction between major and minor phonological phrases that she did in previous work such as Selkirk (1986).

It is possible that there would be no acoustic boundary within this sentence: such an observation suggests either that Φ may restructure more than Nespor & Vogel propose or that Φ is not a domain of final lengthening. The phonological phrase is proposed by Nespor & Vogel, however, as the domain of final lengthening in English: "Syntactic phrasing plays a role in lengthening only in the sense that the phonological phrase is built on the basis of syntactic notions. The reason a distinction [between syntactic and phonological phrases] was not made earlier is most likely due to the fact that the two phrasings often do coincide" (Nespor & Vogel 1986:1990). They relate the blocking of iambic reversal at phonological phrase boundaries to the fact that lengthening in this position increases the interval between adjacent stresses.

Nespor and Vogel do not, however, put forward specific experimental evidence in support of their claim that the phonological phrase is the domain of final lengthening, and the status of the assertion is not clear[17]. As mentioned above, Nespor & Vogel suggest that (unidentified) constituents which are not supported by segmental phonological evidence within a particular language may prove necessary as domains of final lengthening, implying that more than one constituent may show this effect. They allude to prominence relations influencing the relative magnitude of final lengthening between intonational phrases[18], but they make no specific claims for constituents below the level of the phonological phrase, such as the prosodic word. (As discussed in Section 2.2.1, the evidence supporting final lengthening at the word-level is uncompelling, so Nespor & Vogel may be justified in ignoring it.) Thus, the extent to which they regard final lengthening as specifically associated with a phonological phrase domain or as a property of some or all prosodic constituents is uncertain.

If the phonological phrase does manifest final lengthening, it is at first sight sufficient to account for all final lengthening at or above this level: according to the strict layering hypothesis, final edges of higher constituents must necessarily coincide with final edges of lower constituents. The end of every utterance is also the end of an intonational phrase, and the end of every intonational phrase is also the end of a phonological phrase. If lengthening of greater magnitude is observed at the end of the higher constituent, this may be related to the fact that the final phonological phrase in the intonational phrase or the utterance is the stronger. For example, the sentence

There was a large black dog and a tiny brown dog

may be realised as three phonological phrases, with a boundary after the verb and an-

[17]Selkirk (1996:202) also suggests a link between phonological phrases and speech timing: "Independent evidence for or against this right-edge alignment of phonological and syntactic phrases in English will most likely come from phenomena involving intonation or durational patterns, which are argued to be characterised in terms of phrase-sized units of prosodic structure."

[18]Nespor & Vogel (1986:223) say that sentence-final intonation and lengthening seem to indicate that "the last [intonational phrase] of [an utterance] is the strong one."

other before *and*. In this case, Nespor & Vogel would predict final lengthening on the word *dog* in both instances. If the second instance of the word is found to be longer than the first, it is not necessary to postulate intonational-phrase-final or utterance-final lengthening; rather, it is more parsimonious simply to state that the final phonological phrase in the intonational phrase is the strong one and thus shows more final lengthening.

If final lengthening is shown to occur consistently at the end of phonological phrases, there are two possible observations which might require the postulation of lengthening at the end of higher-level domains. Firstly, if the phonological phrase is the only domain of final lengthening, the strong phrase should still show the greatest degree of lengthening even when it is not the rightmost phrase of a set of daughters. (This requires some independent test of phonological phrase strength, which Nespor and Vogel suggest may relate to the location of new information within the dominating intonational phrase.) If, in this case, the final phonological phrase still shows the greatest magnitude of final lengthening, this would suggest that the effect relates to a higher-level domain. Secondly, the locus of the effect may provide information about the possible domain: higher-level constituents may not only show lengthening of greater magnitude but may also show a more extensive locus of lengthening. If final lengthening effects are observed with consistently different loci, this suggests the existence of different domains. A Dutch study by Cambier-Langeveld (2000) examining this is discussed below in relation to the phonological utterance.

Studies which have attempted to find phonological-phrase-final lengthening in English, in the context of hierarchical lengthening effects, are discussed in Chapter 4; the experiment reported there examines the locus of final lengthening in English in relation to utterance-medial and utterance-edge durational effects.

The Intonational Phrase

Although Nespor & Vogel identify the phonological phrase with one particular suprasegmental effect—final lengthening—they do not suggest any particular correspondence with intonational phenomena. Major and minor phonological phrases as defined by Selkirk (1986) are identified as domains of intonational phenomena in other languages: the major phrase is required to characterise tonal phrasing in Xiamen and Shanghai Chinese (Lin 1993); the minor phrase (called the accentual phrase by Beckman & Pierrehumbert 1986) appears to correspond to the domain of occurrence of pitch accent in Japanese. It is not clear, however, whether both of these levels are required in English. As Wightman *et al.* (1992:1709) observe: "In Japanese, [Beckman & Pierrehumbert 1986] find clear evidence for an accentual phrase as a simple grouping of words, but in English, they find the evidence for justifying it much less compelling,

although it is clearly possible to define such a unit."

As its name suggests, the intonational phrase is, in contrast, based upon the domains over which intonation contours extend, where a complete intonational contour includes, at least, a nuclear pitch accent, a phrase tone and a boundary tone, with optional prenuclear accents (see, for example: Ladd 1996). Nespor & Vogel also identify the ends of intonational phrases with the position of potential rule-governed pauses in English, although Hayes (1989) says that the utterance is the domain of non-hesitation pauses. The syntactic information used to define IPs is very general, although certain constructions form obligatory intonational phrases, including parentheticals, nonrestrictive relative clauses, tag questions, vocatives and expletives (Nespor & Vogel say that these are all strings that are in some way external to the root sentence). Beyond this, any number of phonological phrases within the same utterance may be dominated by a single intonational phrase, or a number of intonational phrases may each dominate smaller groups of phonological phrases: "The use of n-ary branching trees avoids the problems [associated with] binary branching trees. That is, since there is no additional constituent structure between [IP] and Φ, restructuring may group any sequence of Φs into smaller [IPs], as long as the division respects the syntactic and argument structure conditions as well as the general timing conditions..." (Nespor & Vogel 1986:205). These timing conditions are identifiable with the factors discussed in Section 2.2: the length of phrases, speech rate and style; additionally, certain "semantic prominence relations" may also affect restructuring at this level.

Some researchers have questioned the inclusion of the intonational phrase within prosodic hierarchies largely motivated by segmental phonological phenomena. Selkirk (1990:195) refers to it as an "honorary constituent of syntax", because it is "more closely tied in to the semantic properties of sentences" than other prosodic constituents, and states that she is "reluctant to accept the existence of a prosodic structure in general solely on the basis of the existence of intonational phrases." Gussenhoven (1992) goes further and suggests that intonational constituents—which he calls "association domains"—actually form a distinct class from prosodic constituents, and do not consistently map to any particular prosodic constituent. A relevant question is whether Nespor & Vogel are correct to make the connection, as discussed above, between segmental phonological effects and suprasegmental "phonetic" effects at a lower level, by identifying the phonological phrase as the domain of final lengthening. If this is justified, then given that there is no doubt that the ends of intonational phrases—through whatever mechanism—show final lengthening, all that would be required to make a compelling case for including both constituents within the same representation would be to show that the former are exhaustively dominated by the latter—that the boundaries of intonational phrases only occur at the boundaries of phonological phrases.

The Phonological Utterance

The utterance is discussed above in relation to final lengthening at the end of syntactically-defined constituents: the definition of the utterance given there as a pause-delimited and relatively self-contained stretch of speech is also a reasonable working definition of the phonological utterance. In practice, the utterance very often coincides with the syntactic sentence: as defined by Nespor & Vogel, the utterance comprises all the intonational phrases dominated by the highest node in the syntactic tree. Utterance-restructuring is possible, however: two short, temporally-adjacent sentences may comprise a single utterance when there exists "a syntactic relation (ellipsis, anaphora) and/or a positive semantic relation (and, therefore, because) between the [utterances] in question" (Nespor & Vogel 1986:244). This process is optional, depending on style and rate of speech amongst other factors.

As mentioned in Section 2.3.1, Nespor and Vogel say that the utterance is the domain, in certain dialects of English, of r-insertion. This may provide evidence of utterance restructuring between sentences; for example, in:

Let's take the car. I'm exhausted.

an /ɹ/ may be inserted between the two sentences if they comprise a single utterance. Shattuck-Hufnagel & Turk (1996) discuss such observations in relation to flapping of /t/ in American English, proposed by Nespor & Vogel to be another utterance-span effect. They say that information is lacking about the intonational contours of utterances, which makes it difficult to identify their occurrence. Between-sentence effects such as flapping, and by the same argument r-insertion, may be a result of two closely-linked sentences being realised as one intonational phrase and thus associated with that domain rather than the utterance.

In experimental investigation of speech timing, sentences are very often spoken in isolation, and by definition comprise a single utterance. Thus sentence-final lengthening may be readily equated with utterance-final lengthening. The problem of identifying final lengthening effects with a particular domain is discussed above in relation to the phonological phrase; here it may be asked whether sentence-final lengthening actually relates to a phonological constituent identifiable as the utterance, or to subordinate constituents such as the intonational phrase or the phonological phrase. As stated above, the locus of lengthening at the end of different constituents may help to resolve this question. Cambier-Langeveld (2000) does not find strong evidence in Dutch of lengthening at the end of phonological phrases, or at least, no greater lengthening than at the end of prosodic words. She does, however, find final lengthening at the end of intonational phrases and utterances, where the locus of the effect is consistent between the two types of boundary: the final syllable rhyme, plus the

penultimate syllable rhyme only in the case that the final syllable contains a reduced vowel. For their part, Nespor & Vogel suggest that patterns of sentence-final intonation and lengthening indicate that the last intonational phrase of an utterance is the strong one, but what these distinctive suprasegmental patterns are is not specified. As discussed in Section 2.2.1, variations in final lengthening between domains in English speech remain uncertain; this issue is discussed further in Chapter 4 and Experiment 2 looks for evidence of final lengthening at utterance-medial and utterance-final word boundaries.

2.4 Lexical stress

2.4.1 The concept of speech rhythm

As mentioned in Chapter 1, the term "rhythm" is used here to mean the pattern of alternation of strong syllables (carrying primary or secondary lexical stress) and weak (unstressed) syllables. The concept of "speech rhythm", whether under this restricted definition or when more widely applied to speech timing phenomena, is both prevalent and elusive. It worthwhile considering briefly what it means to assert that speech is a rhythmical activity, before examining the durational consequences of such an assertion.

Rhythm implies the repeated occurrence of events marked by common features. It is an essentially temporal phenomenon, but not exclusively auditory. Visually-observed movement may be said to possess rhythm through the repetition of some sequence of events in time: the sight of a person walking or running and of the wings of a bird in flight, convey a clear sense of rhythm. Visually-observed rhythms in many instances have a particular quality of flow; that is, a sequence of events may be seen to recur without there being any sense of a privileged point in the sequence. There is no position which may be described as the onset, although one may be arbitrarily assigned. This is similarly true from the point of view of the person observed at a rhythmic activity: one may run, with a greater or lesser degree of rhythm experienced proprioceptively, without any awareness of a point in the sequence of movements which mark the beginning of the next repetition. In other cases, rhythm may arise from a sequence of manifestly discrete events, each with a clear onset in time: a heartbeat proprioceptively experienced is rhythmic and discrete.

In the auditory domain, rhythm seems almost to require this quality of discreteness: it is difficult, if not impossible, introspectively to characterise an auditorily-experienced rhythm without making reference to the onset, or some other specific point, of each event in a sequence. Below a certain rate of occurrence, each repetition of a regular acoustic event is experienced as an equal contributor to the sequence. A

heartbeat heard at a normal resting frequency is a discrete sequence of equally salient or prominent thuds. Above a certain frequency however, the auditory system imposes a structure on a sequence of acoustic events such that the events are perceived as discrete groups, with the first in each group possessing the greatest degree of salience. A very fast heartbeat may be heard as a sequence of paired beats, with the first in each pair being the more prominent.

This grouping tendency, and the conferring of salience on the first event in a group, is clearly a particular feature of the auditory information processing system, discussed by Woodrow (1951:1232):

> By rhythm, in the psychological sense, is meant the perception of a series of stimuli as a series of groups. The successive groups are ordinarily of similar pattern and experienced as repetitive. Each group is perceived as a whole and therefore has a length lying within the psychological present.

The size and perceptual experience of these groups depends to some extent on the rate at which successive stimuli occur. When acoustic events occur in rapid succession, it would seem that the brain devotes most of its processing power to a certain fraction of the events, every second or third or fourth occurrence, the others being perceived largely as inferred and degraded copies of the salient instances. There is an indisputable sense in which the auditory sensation of the prominent events is different from that of the others: the sound of a rapidly ringing bell creates the ineluctable percept that certain of the chimes are different from the others, and introspection is required to demonstrate to oneself that they must in fact be identical. Auditory sampling is reflected in onomatopoeic descriptions of regular acoustic events: the chime of a slowly tolling bell is either *ding* or *dong* according to pitch; at a faster rate it is *ding-dong*; the chime of a rapid alarm bell is *ding-a-ling-a-ling*. Auditory sampling clearly has an upper limit as well: above a certain rate of occurrence, processing is no longer devoted to individual events and the sound is experienced as a continuum, although at relatively low frequencies individual events may be inferred if not perceptually isolated.

This sampling property of the human auditory processing system seems to have influenced the evolution and development of acoustic production systems. The concept of prominence-headed groups of sounds is clearly fundamental in music. Furthermore, the regularity of occurrence of these prominences and disruptions to this regularity are planned by the composer and can be interpreted by the listener. Similarly in speech, certain parts of an utterance are more prominent than others, with salience conferred in production and, it seems likely, in perception as well. Given the existence of these patterns of prominence, it is tempting to believe that, as in music, prominences are produced at regular intervals of time, and disruption from this regularity can be interpreted by the listener. The empirical evidence for such regularity

is lacking, however. In particular, the rhythmical foot, comprising a initial syllable of high prominence and a number of subsequent less prominent syllables has often been asserted as a phonological unit with consequences for the timing of speech production (for example: Lehiste 1977). The evidence reviewed below indicates that, as traditionally defined, there is remarkably little support for the concept of the foot that cannot be better accounted for by reference to other linguistically-motivated structures.

Different conceptions of the metrical foot are examined in Section 2.4.2; durational evidence for the foot as a linguistic unit is discussed in Section 2.4.3; the evidence for a localised durational effect of prominence is presented in Section 2.4.4; the non-durational evidence for the rhythmical foot as a linguistic unit is considered in Section 2.4.5.

2.4.2 Isochrony and the rhythmical foot

Prominence-headed constituents

There are at least as many conceptions of prominence-headed constituents as there are postulated degrees of prominence. Firstly, strong syllables may be defined as those that contain a full vowel and carry some degree of lexical stress, and weak syllables as unstressed and containing a reduced vowel. By these definitions, the most basic prominence constituent into which an utterance can be (almost) exhaustively divided is the cross-word foot: this begins at the onset of a strong syllable and contains all subsequent weak syllables up to the onset of the next strong syllable[19]. Where an utterance begins with an anacrusis—one or more weak syllables before the first strong syllable—under this definition those syllables are extra-metrical, that is, not contained within any cross-word foot.[20]

The cross-word foot so defined corresponds closely to poetical concepts of feet as composed of strong and weak syllables, in most cases one strong syllable and a number of weak syllables. Poetical feet may be left-headed or right-headed and are classified according to the number of syllables they contain: for instance, trochees and dactyls are left-headed disyllabic and trisyllabic feet; iambs and anapaests are right-headed disyllabic and trisyllabic feet. The types of poetical feet and their use in verse are essentially matters of convention, however: to postulate both left-headed and right-headed cross-word feet in the description of natural speech would seem to confound description of an utterance's rhythmical structure rather than assist it. A single word may be described as an iamb or an anapaest, but in describing an utterance,

[19]There are other prominence-delimited constituents which have been called cross-word feet, as discussed below.

[20]Some metrical theories posit silent stresses to deal with anacrusis. For instance, in Abercrombie's (1965) account of the phonetics of verse structure, there are held to be silent stresses as the head of feet before line-initial—but not stanza-initial—unstressed syllables.

it suffices to identify the strong syllables and count the number of unstressed syllables headed by each. Most researchers choose to use a left-headed foot in metrical description, sometimes by convention. Left-headed feet have also been hypothesised to have specific consequences for timing, as discussed below.

The cross-word foot may be regarded as the most basic rhythmical unit, as it reflects the lowest level of prominence contrasts (leaving aside the difference between primary and secondary lexical stress, which, as discussed in Chapter 1, does not seem to have significant timing consequences). Thus, the cross-word foot so defined is referred to here as simply the foot, except where clarification is required, with the number of syllables it contains specified where appropriate. There is another proposed constituent of speech, the within-word foot, also composed of a strong syllables and a number of subsequent weak syllables; as its name implies does not extend beyond word boundaries. It is fundamentally non-exhaustive, as it excludes word-initial unstressed syllables and most function words, and as such is not appropriate for a complete description of speech rhythm. It has been considered as the locus of durational effects such as final lengthening and accentual lengthening, as discussed elsewhere in this chapter.

Certain strong syllables in speech are pitch accented, and perceived as being more prominent than other strong syllables by virtue of that fact. The specific durational consequences of pitch accent are considered in Section 2.5, but is appropriate here to consider the possible existence of a prominence constituent headed by a pitch-accented syllable. This may be defined as beginning with the onset of a pitch-accented syllable and continuing until the onset of the next pitch-accented syllable, encompassing all intervening syllables. This unit—another type of cross-word foot—is sometimes referred to as the Abercrombian foot (after David Abercrombie; for example: Abercrombie 1965[21]). As in the case of the cross-word foot, there may be extra-metrical syllables before the onset of the first Abercrombian foot in an utterance. A prominence-headed constituent may also be postulated from the onset of one nuclear accented syllable up to the onset of the next; the durational evidence for this will be discussed briefly below.

The isochrony hypothesis

All the prominence-headed constituents discussed above have at some time been considered as potentially isochronous units in speech. Early formulations of isochrony, such as that by Classe (1939), are applied by Pike (1945) to a typological distinction between languages such as English, called "stress-timed", and languages such as French,

[21] An alternative interpretation of Abercrombie's formulation is cross-word feet headed by primary lexical stresses. Clearly, primary lexical stress is often, but not always, accompanied by pitch accent.

called "syllable-timed". In stress-timed languages, it is asserted, lexically-stressed syllables occur at regular intervals of time; by contrast, in syllable-timed languages, syllables themselves occur at equal time intervals. Thus, in English the durations of cross-word feet will tend to be equal, however many syllables they contain.

A formal statement of the isochrony hypothesis could be:

Rhythmical feet are of equal duration at constant speech rate.

The mechanism Pike proposes for the timing adjustment required to compensate for the variation in the number of syllables within feet is a compression of unstressed syllables in proportion to their number. Others consider the effect in terms of the shortening of a lexically-stressed syllable as unstressed syllables are added (for example: Jones 1942–43). According to either mechanism, isochrony is a domain-span compression process: an inverse relation between the number of syllables in a foot and the duration of constituents of the foot, independent of their position. Attempts have been made to demonstrate isochrony for different types of rhythmical feet, and as outlined below, have shown the hypothesis to be false.

2.4.3 Experimental studies of isochrony

In the widely-cited paper *Isochrony Reconsidered*, Lehiste (1977) presents a review of the evidence for isochrony of rhythmical feet in speech. As discussed above, prominence-headed constituents may be postulated for each distinct level of prominence in speech. The effect on speech timing of a number of these constituents has been experimentally examined.

Lea (1974) is cited by Lehiste as looking at the relationship between the number of unstressed syllables between lexical stresses and the duration of the interstress intervals. Lea does not find evidence for foot-level isochrony: the average interstress interval appears to increase almost linearly with the number of unstressed syllables in the foot. Lehiste also discusses her own studies of interstress intervals, which use "relatively more homogenous material" (Lehiste 1973; Lehiste 1975): sentences comprising four cross-word feet of varying lengths. As in Lea's study, there is a regularity of duration for feet of the same numbers of syllables, but there are marked differences in the duration of feet of different length, even controlling for position of the foot within the sentence. Lehiste indicates one instance in this study where the same speaker has a difference in mean durations of monosyllabic and disyllabic feet of 133 ms, despite the feet occurring in the same sentence-medial position.

Bolinger (1965) measures intervals between pitch accents, corresponding to the length of the Abercrombian foot as defined in Section 2.4.2 above. As cited by Lehiste, he fails to find evidence for isochrony in the timing of this putative constituent. Of

the 53 Abercrombian foot durations measured, 13 are approximately twice as long as shortest such foot in the study. Lehiste also cites the study by Shen & Peterson (1962) of intervals between "primary stresses"—meaning the primary stress in each sentence—effectively corresponding to a prominence constituent headed by nuclear pitch accents. Their materials, containing sentences of varying length, elicit from all speakers very large ranges in the intervals between primary sentence stresses, with no detectable central tendency.

The review of evidence presented by Lehiste makes it clear that prominence-headed constituents are not produced isochronously in English speech. She nevertheless makes the claim that foot-level isochrony exists (that is, isochrony in the cross-word foot headed by lexical stresses), and suggests that deviations from isochrony are not perceived in the majority of cases. When large enough to be perceived, the deviation in foot-level isochrony may interpreted by the listener as evidence of the presence of a syntactic boundary.

Evidence for her assertion that deviations from isochrony are not perceived is presented from an experiment (Lehiste 1975) which is the perceptual analogue of her studies which fails to demonstrate isochrony in production. She plays recorded utterances composed of four cross-word feet to listeners, and finds that they have "considerable difficulty" in identifying which of the feet is the longest or shortest in any sentence. In contrast, the equivalent task using synthesised materials (clicks separated by noise with the interstress intervals of the spoken sentences) is found to be much easier.

Lehiste proposes this finding as evidence of "perceptual isochrony". She says that "if listeners cannot identify the actually longest or shortest measures in spoken English sentences, the measures must seem to them to have equal duration; if you cannot tell them apart they must be alike. Isochrony would then be a perceptual phenomenon" (Lehiste 1977:256). It is not clear that this conclusion can be drawn: it is one thing to find that a task is difficult, and another to prove that its terms of reference have some perceptual reality. The result could equally well be taken to indicate that listeners do not attend to interstress intervals because they are not linguistically important. If the outcome can be taken as support for two contradictory conclusions, their validity must be questioned.

Perceptual isochrony is questioned by Scott *et al.* (1985), who find that both French and English listeners tend to regularise interstress intervals found in speech more than the beats of a simple non-speech stimulus when they imitate the rhythm of the stimulus. If stress-timing were a property of English, and French were syllable-timed, one would expect to find differences between French and English performance on this task, as French speakers should not have a bias towards perceptual isochrony in

speech. Furthermore, both sets of speakers also regularise the beats on a third version of the task, using non-speech materials with the acoustic complexity of speech. This outcome raises "the possibility that subjects are not actually doing anything very interesting at all—that they are simply exhibiting a response bias toward evenly spaced taps when the task becomes difficult" (Scott *et al.* 1985:161).

Lehiste (1977) acknowledges that the difference in performance between speech and non-speech stimuli might be a gradient distinction; she further suggests that the determination of the just-noticeable-difference (JND) for non-speech on her experimental task might indicate a minimum value for the durational difference required between feet for deviations from isochrony to be observed. Using the same design, with four click-delimited intervals of noise, she systematically varies their duration so that three of the four are isochronous and the fourth, variously located, is longer or shorter. Her finding is that the amount of durational variation required to distinguish the non-isochronous foot from the others ranges from 30 ms to 100 ms. This is sometimes cited as evidence that the JND for duration in speech perception ranges from 30–100 ms (for example: Couper-Kuhlen 1986:54); this is not claimed by Lehiste, however, and it is clear from the task described that this conclusion should not be drawn. Firstly, the task uses non-speech stimuli. Because the task is likely to be easier with sounds of less complexity than speech, it could be argued that the JND represents a minimum threshold for speech. Listeners, however, have expectations based on linguistic experience about the relative durations of speech sounds, and may actually be more sensitive to durational variation which is linguistically interpretable: for instance, Klatt & Cooper (1975) suggest JNDs for individual speech sounds vary according to position within syntactic structures such as words and sentences. The durational range for the non-speech task in Lehiste's experiment only relates to extended stretches of sound designed to correspond to feet. There is no strong evidence that listeners actually pay attention to the duration of feet, or even, as discussed in Section 2.4.5, that the concept of foot has any psychological validity at all.

Lehiste contends that listeners do pay attention to variations in foot duration, provided they are above the threshold of perception: "In English, lengthening of interstress intervals is frequently used to signal the presence of a syntactic boundary" (Lehiste 1977:253). She cites O' Malley *et al.* (1973) as suggesting that boundaries in speech are marked by acoustic cues including lengthening, but takes issue with them because they locate "their boundary signals at given points in a linear sequence, without relating them to the general rhythmic structure of the utterances". This comment implies that it is not the precise location of the lengthened segments that is important but the fact of lengthening of the interstress interval. If the domain-and-locus description of systematic durational variation is valid, however, it is precisely the "given

point" of the lengthened segments that is important, rather than differences in foot duration. As Lehiste herself points out, perceptible lengthening of an interstress interval could be used to signal anything and, if it used in speech to signal a syntactic boundary, lengthening alone cannot indicate anything else, at least where the possibility of a boundary exists however, lengthening of the rhyme of a word-final syllable is less ambiguous as a cue, even in the absence of other suprasegmental cues such as fundamental frequency variation, and allows durational variation in a different locus to perform a different function. Lehiste further proposes that polysyllabic shortening and the shortening of consonants in clusters are evidence of isochrony at work, using a foot-span compression process to explain phenomena which might be better accounted for as localised durational adjustments within the word or syllable.

Lehiste puts forward perceptual evidence in support of the view that lengthening of the interstress interval signals a boundary, referring to the study by Lehiste *et al.* (1976) thus: "disambiguation was produced solely by increasing the length of the interstress interval, and the results of our study show that this is indeed a sufficient cue for signalling the presence of a boundary" (Lehiste 1977:262). It could be argued that although such lengthening is *sufficient*, specifically-located lengthening is a better cue. Furthermore, it is far from clear that the study she cites does, in fact, demonstrate that interstress lengthening is a boundary cue. In particular, she says that in Lehiste *et al.* (1976) the "interstress interval was increased by increasing the duration of each sampling period by the same factor; the durational relationships of the segments to each other remained the same." In fact, in the 1976 study, the durational manipulation is not reliably placed in the interstress interval. Lehiste *et al.* identify nine sentences in which the durational manipulations produced a significant response bias; in only one instance is the only durational manipulation a variation in the length of the cross-word foot (underlined):

The president of the university's committee on education policies came.

In five cases, two adjacent feet are lengthened, as in:

The old men and women stayed at home.

In three cases, part of the sentence is lengthened, whilst another part is shortened, such as:

I know more beautiful women than Mary.

where the underlined words were both and alternately lengthened and shortened.

It is possible to reinterpret these results in a variety of ways: for instance, in the *old men and women* example, as a phrasing effect obtained by delimiting a constituent

with a local speech rate variation. The main point here is that the results of Lehiste *et al.* (1976) do not support a theory of perceptual isochrony in which perceived deviations from isochrony indicate the presence of a syntactic boundary (or indeed, a prosodic boundary).

Finally, to suggest that isochrony is nevertheless present as an "underlying tendency" is equally invalid, implying as it does that stressed syllables would be regularly spaced if not for all the other durational factors clouding the picture. Lehiste offers a quotation which expresses this view: "The implication is that speech attempts to be very rhythmic, but fails because of durational constraints" (Barnwell 1971:88). This view does seem insupportable under the weight of subsequent evidence, both the continual failure to observe isochrony and the replication of experiments which demonstrate other durational effects such as phrase-final lengthening, word-initial lengthening and accentual lengthening. Firstly, although the question of the precise characterisation—within, for instance, a domain-and-locus description—of these well-attested patterns remains open, the patterns themselves have been found to be statistically significant under a variety of conditions, in contrast to isochrony. Secondly, stress-adjacent lengthening is a prominence-related durational effect, discussed in the following section, which is be observed even in the presence of another structural effect—final lengthening—indicating that the presence of one type of timing process need not preclude the observation of another. Finally, segmental restrictions on durational variation are very minor in comparison with structural influences; most phones may be produced over wide range of durations, and so segmental variation in the composition of feet cannot be held responsible for failures to observe isochrony.

2.4.4 The stress adjacency effect

The relative prominence of two adjacent syllables has consequences for speech timing. Bolinger (1965) observes that a full vowel syllable followed by another full vowel syllable is longer than when followed by a reduced vowel syllable. The presence of a full vowel in a syllable almost always indicates the presence of (primary or secondary) lexical stress; likewise, a reduced vowel in a syllable is a reliable indicator that the syllable is unstressed. The observation may be restated thus:

The duration of a stressed syllable is greater when
the following syllable is stressed than when it is unstressed.

The effect is in the direction predicted by the isochrony hypothesis, where the interpretation would be that a disyllabic foot is compressed (or the monosyllabic foot expanded) in order to equalise foot durations. Although, as demonstrated in the previous section, there is very little empirical support for the isochrony hypothesis, there

is strong evidence for what may be termed the "stress adjacency effect". As the evidence reviewed now suggests, this seems to be a strictly rhythmical effect, apparently independent of the boundaries of words and higher-level syntactic or prosodic constituents.

The study of Van Lancker *et al.* (1988) is designed to explore the influence of word boundary and pitch accent upon the occurrence of the effect described by Bolinger. Example sentences designed to test the effect of accent include the following, where the measured syllable (underlined) is followed in the first of each pair by another full vowel syllable and in the second by a reduced vowel syllable. These may be designated the full-full and full-reduced conditions respectively.

Accented condition

They rotate both ways... vs ... *They wrote it both ways.*[22]
John cracked the walnuts... vs ... *John corrected the papers*

Unaccented condition

They describe the electron lineup... vs ... *They describe the electron alignment.*

There are a number of potential problems with the design as illustrated by these sentences. Firstly, it is apparent from the description of the experimental design that pitch accent is equated with primary lexical stress in their materials, which raises the possibility that the presence of accent was assumed rather than detected in their recordings. Secondly, the relationship of the measured syllables to syntactic boundaries is not always constant within sentence pairs: if a boundary of some kind is present after the target syllable in one instance—such as after /roʊt/ in *wrote it* but not in *rotate*—the stress adjacency effect may be masked or exaggerated by domain-final lengthening. Thirdly, the immediate phonetic environment of the measured syllable is often quite different within sentence pairs, as acknowledged by Van Lancker *et al.* Finally, in common with many speech timing studies, the durational effects of syllable number within particular constituents presents a potential confound. In the first sentence pair above, the measured syllable is in a disyllabic word in the full-full condition and in a monosyllabic word in the full-reduced condition; polysyllabic shortening would predict a durational change in the opposite direction to stress adjacency. In the

[22]This an American study, hence the first syllable in *rotate* has primary stress. It may be noted that the syllabification implied here by comparison with *wrote it* is contrary to the principle of maximal onset: that is, the authors are presumably intending to compare syllables with the same structure in the two sentences, but /ɹoʊt/, which is clearly a syllable in *wrote it*, may not be a syllable in *rotate*. Consultation with an American speaker suggests that the syllabification is also contrary to the evidence from flapping: the medial /t/ in *rotate* is unlikely to be flapped, unlike the /t/ in *wrote it*.

other sentence pairs, the sentence in the full-reduced condition has an extra syllable: a domain-span effect operating over some constituent which includes these syllables would predict a durational difference in the same direction as the stress adjacency hypothesis. This point is explored further in relation to studies of polysyllabic shortening in Chapter 3.

Although some of these possible confounds are present in many of the sentence comparisons, there is sufficient diversity of materials to give the results some validity, particularly as confounding factors would favour the null hypothesis in some cases: for example, any effect of word-final lengthening or polysyllabic shortening in the *wrote it* vs *rotate* sentence pair would tend to mask the effect predicted by the experimental hypothesis.

Van Lancker *et al.* (1988:343) report that "a target syllable followed by a full syllable is longer than one followed by a reduced syllable, whether that target is accented or unaccented and regardless of word boundaries which may intervene between the two syllables." The target syllable durations shown in Table 2.3 are for the accented and unaccented conditions, pooled across the word boundary conditions (thus, in some cases the measured syllable is followed by a word boundary and in others not).

	Full-full condition	Full-reduced condition
Accented	177	175
Unaccented	155	133

Table 2.3: Reported durations (ms) of the measured syllables in the study by Van Lancker *et al.* (1988). The durations shown here have been converted from mm of spectrogram, the unit reported by the authors.

Comparison of the two rows of the table shows the lengthening effect of pitch accent, discussed in Section 2.5 below; given that many of the "unaccented" syllables are likely to be found in words containing a previous pitch accent, it is probable that these will also be subject to some degree of lengthening, as suggested by studies such as Turk & White (1999), so the magnitude of the effect may be underestimated here. Comparison of the two columns of the table shows the effect of the following syllable type upon measured syllable duration: it is immediately apparent that the magnitude of stress-adjacent lengthening for unaccented syllables is much greater than that observed for the accented syllables, with 17% lengthening in the former condition and only 1% lengthening in the latter. There is, however, no suggestion in the reported statistical analysis of an interaction between the accent condition and the syllable type condition. Unless due to a misprinted figure, this difference—even if not significant in this analysis—seems worthy of comment.

The findings of Rakerd *et al.* (1987) agree with those of Van Lancker *et al.* in finding that the stress adjacency effect is not blocked by word boundaries. This study goes further in looking systematically at the influence of larger boundaries, with the effect examined across major syntactic boundaries (MSB, corresponding to a noun phrase/verb phrase boundary) and within the same noun phrase (WPH). Rakerd *et al.* record subjects reading aloud sentence pairs such as:

WPH *The strong peach light was unpleasant.* vs *The strong peach delight was unpleasant.*

MSB *The young duke armed his subjects.* vs *The young duke disarmed his subjects.*

The measured syllable (underlined) is a stressed syllable followed by either another stressed syllable or an unstressed syllable; between the two syllables there is either a minor or a major syntactic boundary. Rakerd *et al.* report substantial final lengthening in preboundary syllables, suggesting that the major syntactic boundaries are prosodically realised in many cases, but there are some potential problems with the experimental design as outlined. Firstly, as in the study by Van Lancker *et al.*, the phonetic environment of the measured syllable is not kept constant: segmental effects on duration are likely to be small and unsystematic, however. Secondly and more importantly, there is no control of the distribution of pitch accents and no account given of their occurrence: inspection of the materials suggests that the measured syllables are likely targets for pitch accent in many cases.

The results show that "shortening of a stressed target syllable by an unstressed syllable across a word boundary was not significantly less at an NP/VP boundary than elsewhere, even though the target word shows substantial domain-final lengthening at the boundary" (Rakerd *et al.* 1987:152), supporting and extending the findings of Van Lancker *et al.* It may be noted, however, that the observed stress-adjacent lengthening is very small: a mean of 7ms, where the duration of the measured syllable in the non-adjacent condition was in the range 135ms–190ms. In comparison, Van Lancker *et al.* find a mean of 22ms stress-adjacent lengthening in their unaccented condition, with a non-adjacent mean duration of 133ms.

Taking both studies into account, the simplest explanation of the smaller stress adjacency effect observed by Rakerd *et al.* is that the measured syllables in their materials are, in many cases, pitch accented; furthermore, that the attenuation of stress-adjacent lengthening in the presence of accentual lengthening suggested by Van Lancker *et al.*'s data is a real effect. This conclusion is, of course, speculative without access to the original recordings, so the hypothesised interaction between stress adjacency and pitch accent remains unproven at present.

Fant *et al.* (1991) present the stress adjacency effect as the sole major timing consequence of the rhythmical organisation of speech, summarising the situation thus:

> A foot's duration increases on the whole proportionally to the number of unstressed phonemes or syllables contained within it. The increment per segment is close to the average duration of unstressed segments. Foot shortening effects as a trend towards isochrony have been much discussed in the literature [...] The main effect appears to be in the step from none to one following unstressed syllables in the foot. However, in our experience, these effects are marginal and not sufficient as a basis for a theory of "stress timing" (Fant *et al.* 1991:84).

If this analysis is correct, then one may ask whether the concept of the foot is required to characterise the stress adjacency effect. Fowler (1990) suggests that it might be, saying that "the lengthening of a stressed syllable at the right edge of a foot [...] may count as a sort of preboundary lengthening." Likewise, Rakerd *et al.* (1987) investigate what they call "foot-level shortening", but the experimental evidence they provide relates only to the monosyllabic vs disyllabic foot effect highlighted by Fant *et al.* (1991).

An alternative characterisation would eliminate the foot altogether and describe the effect as arising purely from the adjacency of stressed syllables. In order to distinguish this "stress adjacency" account from the "foot-level" account, certain predictions might be made. Firstly, only the stress adjacency effect could predict any lengthening of the second of two adjacent stressed syllables. Secondly, only the foot-level effect could predict any shortening in the disyllabic vs trisyllabic foot comparison, for instance, in the case that the effects of foot-finality diminished gradually with distance from the end of the foot.

As seen above, Fant *et al.* (1991) suggest that the latter effect is not observed. In contrast, Rakerd *et al.* (1987:154) cite Fowler (1981) as providing evidence that "Two unstressed syllables shorten a stressed syllable more than does one", and furthermore that the shortening effect is confined (or largely confined) to the first stressed syllable. Examination of the method used in Fowler's study shows, however, that it cannot be used to support the assertion made by Rakerd *et al.* The experiment uses reiterant words composed of stressed syllables (/si/, /sʌ/) and unstressed syllables (/sʌ/) in a fixed frame sentence. The durational effect on stressed syllables of the number of preceding and following unstressed syllables is reported. The duration of stressed syllables is found to be inversely related to the number of unstressed syllables which follow it, an effect Fowler calls "anticipatory shortening"; limited evidence is found for an inverse relationship between stressed syllable duration and the number of unstressed syllables preceding it (a hypothesised effect Fowler calls "backward shortening"). The problem with the interpretation of these results is that the recording procedure requires subjects to produce the sentence "in time" with a metronome; that is, stressed syllables are placed on regularly-occurring beats. The procedure is used so that the coarticulatory correlates of a large and reliable anticipatory lengthening ef-

fect can be examined, and Fowler notes that "metronome pacing may strongly inflate the anticipatory shortening parameter and may even decrease the backward shortening parameter" (Fowler 1981:49). Given this, Rakerd *et al.* are mistaken in drawing any conclusions from this experiment about the durational effect on the second of two stressed syllables of intervening unstressed syllables, and about the consequences for both stressed syllables of more than one unstressed syllable intervening between them.

The experimental design used by Fourakis & Monahan (1988) apparently allows direct consideration of these questions. They measure the duration of pairs of stressed syllables when separated by either one or two unstressed syllables. The sentences are constructed with the template:

(your) machines will (soon) pass (at) my desk today

where the measured stressed syllables are underlined and the presence of the bracketed words is systematically varied.

At first sight, the results seem contradictory, offering apparent support for both the foot-level hypothesis and the stress-adjacent hypothesis: for the former because there is a shortening effect in the disyllabic vs trisyllabic foot context; for the latter because this shortening affects the following stressed syllable as well as the preceding stressed syllable.

There are, however, a number of confounding factors which may contribute to this puzzling state of affairs. Firstly, the presence or absence of additional stressed syllables affects not only foot structure but the relationship of the measured syllables to phrase boundaries—and may affect the placement of those boundaries—as well as the length of the utterance and other constituents. Secondly, lexical stress may once again be confounded with pitch accent: at least some of the measured stressed syllables are likely to be pitch accented. Indeed, it is possible that pitch accents are what Fourakis & Monahan intend to delimit their feet: certainly, words such as *will* and *soon* are not always lexically unstressed.

Finally deciding between the competing stress adjacency and foot-level hypotheses is beyond the scope of this dissertation, although some of the possible confounding factors in previous studies, such as the effect of utterance length on stressed syllable duration, are addressed experimentally. What seems clear, however, is that the adjacency of lexically-stressed syllables lengthens the first syllable compared with when one or more unstressed syllables are interposed. It seems a strong possibility that this stress adjacency effect is greater when the first syllable is unaccented. There is little unambiguous evidence for the application of foot-based timing rules beyond this localised effect however.

It appears rather awkward to define the stress adjacency effect in the domain-and-locus terms that may be applied to durational effects within syntactic or prosodic constituents. Perhaps the most precise characterisation would be to say that the domain is the spondee—a foot of two strong syllables—and the locus is the first of these two syllables. Such a domain would seem to belong to a separate representation of speech timing from the constituent-based structures discussed in Sections 2.2 and 2.3. Indeed, as suggested by Rakerd *et al.* (1987), the stress adjacency effect appears to operate across constituent boundaries and independently of the timing effect of these boundaries.

2.4.5 The rhythmical foot as a linguistic unit

Given the lack of evidence for the timing consequences of rhythmical foot structure, it is worth considering briefly whether there is any other evidence for the foot's linguistic existence: that is, do rhythmical feet constitute part of the grammar of English, or are they merely a descriptive tool best applied to restricted domains such as poetics?

Within the area of segmental phonology, there does not seem to be evidence of processes that require a domain delimited by the boundaries of the cross-word foot. Nespor & Vogel (1986) describe certain rules of English phonology as operating in a within-word foot domain, but they argue that segmental processes are regulated by a prosodic structure that is partially determined by syntax. It seems unlikely that a constituent which by definition ignores syntactic boundaries, such as the cross-word foot, could be accommodated within a framework utilising prosodic constituents, which are constructed with specific reference to syntax.

With regard to suprasegmental phenomena, within-word feet—but not cross-word feet—have been proposed as relevant to the description of final lengthening, but recent evidence suggests the locus of the effect may sometimes extend over more than one within-word foot (see Section 2.2.1). In intonational research, recent work on tonal alignment in Greek, Dutch and English, such as Arvaniti *et al.* (1998) and Ladd *et al.* (1999), has indicated the primary importance of the structure of the accented syllable—rather than larger constituents such as the foot or the word—in determining the location of intonation peaks and valleys. Rhythmical constituents do not appear to be relevant, although more research remains to be undertaken in this area[23].

[23]In studies carried out so far, the design of experiments has been influenced by the need to have adjacent pitch accents some distance apart, to avoid distortion of the intonation contour due to "tonal crowding". It would be interesting to vary the degree of tonal crowding in such experiments, in order to examine the hypothesis that there is some correlation between the distance from one accented syllable to the next—this corresponds to the Abercrombian foot, by one definition—and the position of the accent maxima and minima. An alternative hypothesis would be that tonal crowding only has consequences for alignment under conditions of adjacency; that is, when one pitch-accented syllable immediately follows another.

Although segmental and suprasegmental evidence for the rhythmical foot is lacking, there may be some usefulness in conceptions of prominence-headed constituents to describe the *distribution* of lexical stresses in English speech. In the prescriptive domain of poetic verse, for which conceptions of rhythmical feet were originated, there is rule-based alternation of strong and weak syllables and the choice of lexical items is in part conditioned by the demands of the particular foot structure of the type of verse. Furthermore, there are constraints upon the number of feet per line. Abercrombie (1965) alludes to this when he says that prose (and by implication, speech) is rhythmic but not metrical, but verse is both rhythmic and metrical: prose exhibits organisation at the level of the foot, whereas verse is organised in terms of feet, and also in terms of a superordinate metrical unit, the line. This is evident, for instance, in verse of iambic pentameters, such as[24]:

> *Was this the face that launched a thousand ships,*
> *And burnt the topless towers of Illium?*

in which the choice of words has been influenced by the requirement to alternate weak and strong syllables so as to form iambuses—right-headed disyllabic feet—and furthermore by the requirement that there be five such feet per line.

Abercrombie's assertion that prose (and also speech) are not metrical in the sense that he intends for poetry seems uncontroversial. Neither syntactic nor prosodic constituents require a certain number of feet: for example, noun phrases, verb phrases, and intonational phrases may all be well-formed as monosyllables. Syntactic phrases furthermore have no theoretical upper limit on their length. Prosodic constituents—especially higher-level constituents such as the intonational phrase—do tend to be limited in length, as outlined in Section 2.1: two small phonological phrases may restructure as a single larger phrase; potentially lengthy intonational phrases may be realised as two smaller phrases of approximately equal size. There is no evidence, however, that these processes make reference to the number of feet within constituents: although Nespor & Vogel (1986:203) refer to rules governing the restructuring of phrases as reflecting "a general abstract rhythmic organisation of speech into chunks of a more or less fixed temporal length", it seems clear that here the term "rhythmic organisation" is being used in the much broader sense of timing in general. The use of the term "abstract" seems questionable in this context, as surely the restructuring of phrases is a matter of speech performance rather than underlying representation, as suggested by their later assertion: "for both [intonational phrase] and [utterance] restructuring, what seems to be crucial is some average length in terms of timing, most probably due ultimately to physiological considerations" (Nespor & Vogel 1986:240).

[24]Christopher Marlowe (1604), *Dr Faustus*, Act 5, Scene 1.

Speech is therefore not metrical in the Abercrombian sense; furthermore, there seems no compelling reason to regard it as rhythmical in the sense that he intends. That is, there is no statistically-based support for the idea that feet themselves have some organisational significance in speech. The rhythmical pattern of speech is very largely determined by the interaction of lexical phonology and syntax: the words in an utterance are chosen on the basis of meaning and not for their pattern of strong and weak syllables[25], and how those words are ordered within a sentence is a matter of syntactic organisation. What arises from these processes has a manifest rhythm in the sense that there are alternations of stressed and unstressed syllables, but this rhythm is almost entirely an emergent property, contingent upon other processes that make no reference to lexical stress.

There are observable regularities in the rhythm that emerges from these lexical and syntactic processes, regularities that are to some extent language-specific. Scott *et al.* (1985) say that English "conspires" to place stresses at regular intervals. At the simplest level, lexical words—containing at least one stress—and unstressed function words tend to alternate: long collocations of function words or lexical words are rare; furthermore, lexical words of several syllables contain secondary stresses. Thus, as observed by Dauer (1983:58), "in a text of conversational English, the vast majority of interstress intervals contain from one to four syllables, with a limit of five [...] Counterexamples are difficult to construct in English." She contrasts this situation with Spanish, which allows only one stressed syllable per word, and where interstress intervals may contain as many as seven syllables.

The maintenance of this emergent rhythmical regularity is occasionally subject to processes which may be seen as entirely rhythmical. The occurrence of a long string of weak syllables often induces the placement of a stress upon one of the syllables, with the stress likely to be towards the centre of the "underlying" interstress interval. The phenomenon of stress retraction may also be seen an purely rhythmical process: when a word-final stressed syllable precedes a word-initial stressed syllable, the primary stress in the first word may be retracted, thus *thir'teen men* may become *'thirteen men*. The existence of this effect is not uncontroversial, however. Cooper & Eady (1986) fail to find acoustic correlates of stress retraction in cases like the *thirteen men* example. Grabe *et al.* (1994) and Grabe & Warren (1995) likewise find that the two stressed syllables in contexts such as *thirteen* are not different in fundamental frequency range, duration and amplitude, except where the second syllable precedes a phrase boundary and/or carries a nuclear accent and is thus marked by greater lengthening and frequency range; furthermore, where the two syllables have equal acoustic prominence,

[25]There are statistical regularities of rhythm within the lexicon, such as the predominance of word-initial stressed syllables (Cutler & Carter 1987), but there is no evidence that word selection takes account of whether these regularities are or are not observed.

listeners cannot reliably judge which is stressed when they are excised from context. In related research, Shattuck-Hufnagel *et al.* (1994) analyse stress shift in terms of the likely placement of pitch accent within a word rather than variation in the acoustic correlates of lexical stress.

In their analysis of stress retraction, Shattuck-Hufnagel *et al.* refer to the tendency for the first pitch accent in a phrase to be placed as early as possible, with the final accent tending to be placed late. In an intonational phrase with, by default, an early accent and a late accent, the length of the phrase may favour the placement of additional accents: long phrases will have three, four or more accents, which will tend to be evenly spaced throughout the phrase, subject to the availability of suitable accent sites. Such an observation could be restated in terms of feet—"Abercrombian feet[26] tend to be of equal size"—but taking this step may imply timing consequences that are not intended in a statement about distribution.

For both lexical stresses and pitch accents, there is no need to invoke prominence-delimited constituents to capture statistical regularity in the occurrence of prominences. The further step of postulating prominence-delimited constituents as the units underlying the control of speech timing in English speech is not only unnecessary, but incorrect, as the evidence discussed in Section 2.4.3 indicates. Domain-span timing processes do not operate over constituents delimited by prominences.

In early speech research, where instrumental evidence was difficult to obtain, rhythmical concepts were arrogated from the prescriptive domain of poetics, but the relevance of such concepts to the science of speech is not supported by empirical techniques now available. Despite this lack of evidence, however, research is still pursued within frameworks that, at least implicitly, presuppose the existence of prominence-delimited constituents. Conceptions of the metrical grid that bracket groups of syllables according to levels of prominence are thus positing a unit which is not empirically motivated. Likewise, the view that "rhythm is manifested as the temporal binding of events to specific and predictable phases of a superordinate cycle" (Cummins & Port 1998:147), is not only implicitly invoking the rhythmical foot as an organisation principle of speech, but is going beyond Abercrombie to suggest that speech, like poetical verse, has an higher level of metrical coordination.

Cummins & Port (1998) cite evidence to support this view, based upon their experimental paradigm of "speech cycling": they require subjects to repeat short phrases, such as *big for a duck*, containing two stressed syllables which are to be aligned with the low and high tones of a metronome for which the timing intervals are varied. They find that English speakers tend to place the onsets of stressed syllables more regularly than the occurrence of the tones: that is, stresses align with points within the over-

[26] Abercrombian feet are sometimes delimited by pitch accents, as in this instance, and elsewhere by primary lexical stresses.

all phrase repetition cycle which divide the cycle into regular intervals. They suggest that the patterns can be understood as the nesting of one unit (the stress foot) within a larger unit (the phrase repetition cycle).

As the authors repeatedly state that the observed co-ordination between stress placement and the higher-level cycle is task-specific, it is unclear why the results should be thought to relate to the performance of normal speech. They state themselves that: "This dynamical interpretation of the speech cycling task suggests why isochrony should be more readily observable within the confines of a repetition task than in normal conversational speech. Repetition generates a stable cycle to which nested processes can (or must) entrain. The changing demands of unconstrained speech production do not allow this stability to persist, though it would emerge occasionally as the speech content permitted. In its simplest form, the speech cycling task provides a stable period [the phrase repetition cycle] within which nested and entrained periods can be seen to emerge" (Cummins & Port 1998:150). They later claim, however, that "By structuring an utterance so that prominences (stresses, beats) lie at privileged phases of a higher-level prosodic unit, rhythm is seen as an organisational principle which has its roots in the coordination of complex action and its effect in the realm of prosodic structure" (Cummins & Port 1998:167).

There appear no grounds for the latter assertion: Cummins & Port admit that the superordinate cycle within which stresses are aligned is a product of the experimental task and make no suggestion about which prosodic constituent might fulfil such a function in normal speech. Moreover, it is one thing to show that subjects are capable of performing a task and quite another to claim that this behaviour is habitual: it would be spurious to infer, from the fact that people can learn to juggle, that the timing of hand movements when eating with a knife and fork is entrained to some superordinate cycle.

Finally, it is interesting to note the lack of intuitive evidence for isochrony even in the rhythmically and metrically-ordered domain of poetic verse. Iambic pentameters should be ideal material for the observation of isochronous performance, with each foot in a regular line being disyllabic. In performance, however, the stresses frequently occur with sufficient irregularity to obscure the fact that verse rather than prose is being spoken. Indeed, a performer attempting to place each stress "on the beat" would sound unnatural and possibly comical.

2.5 Pitch accent

2.5.1 The distribution of pitch accents

All intonational phrases include at least one pitch accent. The primary acoustic correlate of pitch accent occurs in the fundamental frequency contour: commonly, there is a rise in F0, but there may be a fall instead, or even a level stretch corresponding to the absence of normal declination. All types of prominence-lending variation in fundamental frequency may be collectively termed "pitch accents".

Ladd (1996) reviews a phonological classification of pitch accents according to high and low pitch targets, based upon the work of Pierrehumbert and others. He also presents an account of the placement of phrase-level prominence in terms of focus; for current purposes, the important facts are that each pitch accent within a phrase is located on a lexically-stressed syllable (with occasional exceptions due, for example, to contrastive focus) and the final pitch accent—known as the nuclear accent—is perceptually the most prominent. Pitch accents tend to be associated with primary lexical stresses, but tonal crowding may cause movement of prenuclear accents onto secondary lexical stresses (for example: Shattuck-Hufnagel *et al.* 1994).

Although pitch accents are characterised primarily by fundamental frequency variation, they are also associated with significant increases in duration in the vicinity of the pitch accent. Klatt (1976:1210) cites Coker *et al.* (1973) as finding that "[one] acoustic correlate of emphatic or contrastive stress is an increase in the duration of a word by 10%–20% or more." Later researchers such as Sluijter (1995), Turk & Sawusch (1997) and Turk & White (1999) have attempted a more precise characterisation of the distribution of such lengthening. Before reviewing this experimental work, it worth considering again a terminological point raised in Chapter 1. Turk & Sawusch (1996) present evidence relating to what they call the "domain" of accentual lengthening. In the current work, the term "domain" is reserved for the unit which occasions the durational variation—such as the intonational phrase or the utterance in the case of final lengthening—and the term "locus" is used for the unit which actually manifests the effect: in phrase-final lengthening, typically the rhyme of the final syllable. This use of the term "locus" seems to be applicable, at least to a first approximation, to durational processes associated with pitch accent; as discussed below, however, the term "domain", which seems valid and useful in the description of both domain-edge effects and domain-span effects such as polysyllabic shortening, may be rather less susceptible to precise characterisation in relation to pitch accent.

2.5.2 The locus of accentual lengthening

Turk & Sawusch (1997) show that the locus of accentual lengthening extends to all

parts of the accented syllable and beyond. Firstly, they measure the durations of vowels and consonants in the context of near-homophonous word pairs such as *beef arm* and *bee farm*, where the central consonant—/f/ in this case—could be either word-initial or word-final. As expected, the vocalic nuclei of accented syllables show a large amount of lengthening, having a mean duration of 206 ms compared with 163 ms in the unaccented context. The effect on neighbouring consonants is analysed by measuring the duration of the central consonant when word-initial or word-final, in accented or unaccented words: the word-initial consonants show a large lengthening effect due to pitch accent, being approximately 34 ms longer compared with an unaccented mean of about 101 ms; the word-final consonants also show significant lengthening in an accented word, but the magnitude of the effect is smaller, approximately 10 ms greater than the unaccented mean of around 83 ms.

Turk & Sawusch also examine the durational effect of pitch accent upon unstressed syllables adjoining the accented syllable. Again they use pairs of near-homophonous two-word phrases, this time with a central unstressed syllable which could be either word-initial or word-final, for example: *thank fulfil* and *thankful Phil*. They find that word-final unstressed syllables show a large amount of lengthening in an accented word—approximately 20 ms compared with the unaccented context of 150 ms—whereas word-initial unstressed syllables show no significant effect of membership of an accented word. From this they conclude that the locus of accentual lengthening is most likely to be a unit beginning with a lexically-stressed syllable and including, at least, a subsequent unstressed syllable within the same word. This unit possibly corresponds to the within-word foot, although other within-word candidates are not ruled out by the available evidence.

This finding contrasts with that of Sluijter (1995) who examines the durational effects of pitch accent in English disyllabic words. She measures both syllables of words contrasting only in lexical stress placement—such as *ˈcompact* and *comˈpact*—when accented and unaccented. As found by Turk & Sawusch, word-final unstressed syllables show large amounts of accentual lengthening, 38 ms longer than the unaccented mean of 236 ms. Sluijter, however, finds that word-initial unstressed syllables are also lengthened when the word is accented, by 23 ms compared with 147 ms mean duration in the unaccented condition.

Turk & Sawusch consider an alternative explanation of their findings which suggests a possible reason for this discrepancy. In their materials, word-initial unstressed syllables, when in an unaccented word, are preceded by a pitch-accented syllable, albeit with an intervening word boundary. In the accented word condition, word-initial unstressed syllables are followed by an accented syllable without an intervening word boundary. For example,

Unaccented word condition *THANK fulfil*

Accented word condition *thank fulFIL*

where capitals indicate the placement of the accent. If there were a rightward lengthening effect of pitch accent, attenuated but not blocked by a word boundary, which matched a leftward within-word effect in the accented word condition, then this could explain the failure to observe any accentual lengthening on word-initial unstressed syllables.

Turk & White (1999) replicate and extend this experiment by Turk & Sawusch, including a "baseline" condition in which there is no pitch accent on either word in the near-homophonous phrase pairs:

Baseline condition *thank fulfil*

Comparison of word-initial unstressed syllables in this condition with the accented word condition shows that there is indeed a small lengthening effect to the left of the primary stressed syllable within accented words, and also a rightwards lengthening effect attenuated across a word boundary. These results are shown in the top line of Table 2.4. Another experiment reported by Turk & White indicates that this cross-word-boundary effect is not influenced by the strength of syntactic affiliation of the words on either side of the boundary, at least within the context of a verb phrase.

Example phrase type	Unaccented baseline duration	Lengthening: Phrase-initial accent		Phrase-final accent	
thank fulfil	165 ms	7 ms	4%	8 ms	5%
thankful Phil	166 ms	21 ms	13%	4 ms	2%
knee capsize	207 ms	10 ms	5%	7 ms	3%
kneecap size	206 ms	26 ms	13%	-1 ms	0%

Table 2.4: Data taken from Turk and White (1999). The size of the accentual lengthening effect for word-initial and word-final unstressed syllables and secondary stressed syllables compared with the unaccented baseline condition. Example phrase pairs are shown for each context, with the measured syllable underlined.

Notwithstanding the small cross-word-boundary effect to the right of a pitch-accented syllable, further results reported by Turk & White and shown in Table 2.4 strengthen the conclusion that the distribution of accentual lengthening is primarily related to word structure. Firstly, they show that stressed as well as unstressed syllables adjacent to a pitch-accented syllable undergo lengthening: in phrases such as *knee capsize* and *knee capsize* the central syllable, which carries secondary lexical stress, shows a very similar pattern of accentual lengthening to the unstressed syllable in phrases such as

thankful Phil and *thank fulfil*. This result rules out the within-word foot as the locus of accentual lengthening, as a secondary stressed syllable is separately footed from the primary stressed syllable within a word. Secondly, Turk & White find lengthening on both syllables which follow the pitch-accented antepenultimate syllable in words such as *property* and *alternative*. The magnitude of lengthening, is greatest on the pitch-accented syllable itself (30 ms, 23 %), and is greater on the word-final syllable (22 ms, 14%) than on the penultimate syllable (12 ms, 11%), although direct comparisons are only suggestive as the phonetic composition of the syllables is different in each case, and the penultimate syllables are substantially shorter even in the unaccented comparison.

Turk & White (1999:171) conclude that "both the left edge of a pitch accented syllable and the left and right edges of a word-sized unit thus appear to attenuate the spread of accentual lengthening." It is interesting to note the contrast between these results and reported studies of domain-edge effects such as initial lengthening and final lengthening. In the latter, for the most part, the effect appears to be well constrained within the boundaries of its manifest locus. For accentual lengthening, however, no single unit fully characterises the locus, because lesser degrees of lengthening extend beyond its apparent boundaries. The best available description of the locus would be "a unit beginning with a primary stressed syllable and extending rightwards to a word boundary", with additional small amounts of lengthening on syllables adjacent to the primary stress, rightwards across a word boundary but leftwards only within a word. An alternative possibility, particularly suggested by the results of the Turk & White study of trisyllabic words, is that the locus of accentual lengthening is actually bimodal: firstly, there is a large lengthening effect on the accented syllable itself, which is attenuated but not blocked by syllable-initial and syllable-final boundaries, causing a small amount of lengthening on adjacent syllables; secondly, there is a fairly large lengthening effect on the final syllable of the accented word. Where the word-final syllable is adjacent to the accented syllable, as in *THANKful*, the accented-word-final effect combines with a smaller lengthening effect due to adjacency to the accented syllable. Where the syllable is word-initial and thus precedes the accented syllable, as in *fulFIL*, there is only the small amount of lengthening due to the adjacency effect. One of the purposes of the experiment presented in Chapter 4 is to determine whether the bimodal locus or the rightward locus or some other interpretation best characterises the distribution of accentual lengthening.

2.5.3 The domain of accentual lengthening

The findings described in Section 2.4 about the timing consequences of lexical stress lend support to the notion that the constituents that undergo domain-edge durational

processes arise from a different type of representation from that in which the timing effects of prominence might be derived (an issue discussed in more detail with regard to Dutch by Sluijter & van Heuven 1995). For pitch accent, however, constituent boundaries appear to exert some influence on the precise distribution of its durational effects. The interaction between accentual lengthening and syntactic/prosodic constituents, particularly the lexical word, will be considered in the experiments presented in Chapters 3 and 4.

A related issue which remains unresolved is the place of accentual lengthening within a domain-and-locus description of speech timing. As just seen, the concept of the locus seems less clear-cut than in discussions of domain-edge effects. It is even less clear how the domain of accentual lengthening might be characterised. The natural answer would be a constituent which contains a single pitch-accented syllable, but there does not appear to be much evidence for this unit in English, where even a single word may in some cases contain more than one pitch accent: Beckman & Pierrehumbert (1986) propose such a constituent—the accentual phrase—but this appears to be better motivated for Japanese than for English. Alternatively, if accentual lengthening was only associated with nuclear pitch accent, then its domain would be the intonational phrase. Studies of the durational effects of accent have tended to elicit it through contrast (for example: *I said "PROPERTY sale", not "LEGACY sale"*), where the target accent is intended to be the last or only accent in the intonational phrase. The durational effects of prenuclear pitch accents, and of accents not elicited contrastively, have not been much studied. Given the parallels between prenuclear and nuclear accents in terms of peak alignment discussed by Silverman & Pierrehumbert (1990), and also the apparent ubiquity of durational phenomena in relation to other aspects of prosodic and metrical structure, it seems unlikely however that prenuclear accents should not show durational correlates similar to those seen for nuclear accents. Thus, there is not an obvious candidate for the domain of accentual lengthening within a constituent hierarchy.

It remains an open question whether attempts to unify prosodic constituents and prominences within a single type of representation will ultimately succeed. The durational evidence presented in Section 2.4 suggests that the effects of lexical stress in particular may require a separate type of representation; for accent, however, the association of the locus of lengthening with word structure suggests an accommodation within a constituent hierarchy, although the concept of the domain appears less useful for accentual lengthening than for processes at constituent boundaries.

2.6 Summary

In this chapter, the evidence for suprasyllabic timing effects in English speech is reviewed and placed, wherever possible, into a domain-and-locus framework, which incorporates domain-span processes, where the length of some constituent and duration of some subconstituent are inversely related, and domain-edge processes, where the duration of subconstituents is greater at the initial or final edge of a domain than domain-medially. Another durational process, accentual lengthening, is less straightforwardly accommodated within the domain-and-locus framework; furthermore, the durational effects of pitch accent appear often to be overlooked in previous research relating to domain-edge and domain-span processes.

Possible domain-span processes considered include foot-level isochrony, utterance-level shortening and polysyllabic shortening. The available evidence strongly suggests that foot-level isochrony does not exist in English speech. There is, however, strong evidence for a localised rhythmical process, the stress adjacency effect, which appears to operate in a domain which is distinct from the constituents defining the occurrence of domain-edge effects such as final lengthening and initial lengthening. The stress adjacency effect has to be controlled in experiments relating to other durational processes, such as those presented in Chapter 3 and Chapter 4.

There is insufficient evidence available regarding the other domain-span processes. It is possible that utterance-span shortening reflects a general tendency for domain-span processes to operate within all prosodic domains; alternatively, reported results may reflect indirect effects of constituent length, such as restructuring. Experiment 2 in Chapter 4 is, in part, designed to test claims about utterance-span shortening. The discussion of polysyllabic shortening indicates that some of the previous experiments relating to this effect have contained confounding factors such as stress-adjacent lengthening, phrase-final lengthening and accentual lengthening. The design of these experiments is examined in detail in Chapter 3; Experiment 1 is designed to eliminate some of these potential confounds. Word-level durational processes and their interactions are further examined in Experiment 2 presented in Chapter 4.

The existence of domain-edge processes is better attested, although the precise nature of the domains involved is not always clear. Final lengthening occurs at the end of utterances and at the end of certain utterance-medial phrases: whether these observations reflect a single underlying domain or a hierarchy of domains is uncertain. The evidence for initial lengthening indicates quite strongly that it relates to a word-type domain, which may be the lexical word or the prosodic word. An apparent interaction between initial lengthening and lexical stress or pitch accent has been observed, as has a possible utterance-initial shortening effect. The design of experiments which suggest hierarchical domain-edge effects is critically evaluated in Chapter 4; differences

between utterance-edge and utterance-medial processes are addressed in Experiment 2, particularly with regard to the locus of the effect in each case, and to its interaction with the durational effects of pitch accent.

There are a number of different questions to be addressed by the experimental work described in this dissertation. These questions have been raised in this chapter and are presented in the following chapters as formal experimental hypotheses. Testing these hypotheses will contribute to answering three general questions:

- Are previously-observed domain-edge and domain-span effects specific to certain constituents, or do they reflect tendencies at all levels?
- Does each domain-edge and domain-span process have a phonologically-defined locus? Does the locus vary between different levels of hierarchical processes?
- Is the presence of pitch accent a prerequisite for the occurrence of certain durational effects, or does accentual lengthening simply serve to amplify existing effects and possibly modify their loci?

Experiment 1, presented in Chapter 3, looks for evidence of word-level durational processes. Experiment 2, presented in Chapter 4, is designed to determine the *type* of processes—whether domain-edge or domain-span—which operate at the word level and the utterance level.